

7-7-2012

Research Data Curation, Discovery, and Dissemination

D. Scott Brandt

Purdue University Libraries, techman@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

Brandt, D. Scott, "Research Data Curation, Discovery, and Dissemination" (2012). *Libraries Faculty and Staff Presentations*. Paper 6. http://docs.lib.purdue.edu/lib_fspres/6

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.



July 13, 2012

E-Science Outreach & Learning: University of Oklahoma School of Library and Information Studies

**Research data curation,
discovery, and dissemination**

Agenda

- Background: struggling with e-Science
- Data Curation Profiles: who's willing to share what with whom, and when?
- But where are you gonna put this data???
(Purdue University Research Repository)
- Collaborations needed to make this all work
- Data services approach (“data deeds & they’re done dirt cheap”)

Setting the stage

2004: Purdue Interdisciplinary Research Initiative revealed many data needs on campus



What faculty said...

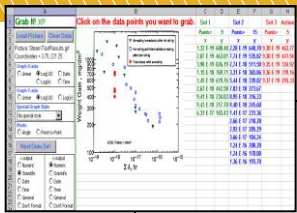
- Not sure how or whether to share data
- Lack of time to organize data sets
- Need help describing data for discovery
- Want to find new ways to manage data
- Need help archiving data sets/collections

Define curation

- **Curation** is the activity of managing and promoting *the use of data, starting from the point of creation*, to ensure its fitness for contemporary purposes and availability for discovery and re-use.
 - **Archiving** is a curation activity which ensures that data is properly selected and stored, can be easily *accessed* and that its logical and physical integrity is maintained over time.
 - **Preservation** is an archiving activity in which specific items of data are *maintained* over time so that they can still be accessed and understood through succession and obsolescence of technologies.

“...active management and enhancement of digital information assets for current and future use.”

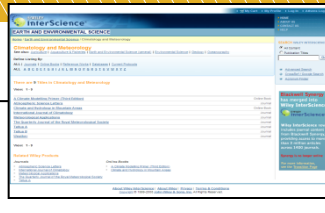
- ✓ active as opposed to passive...
- ✓ enhancement includes metadata to find & use
- ✓ discoverable by and available to wider audiences over longer periods of time



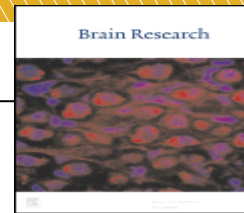
Data repository



IR



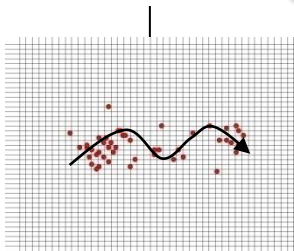
online research
Internet, OA, etc.



published research
traditional



secondary/tertiary resources



Librarians provide education on depositing and finding data.

Year	Country	Year2	Country2	Year3	Country3	Year4	Country4	Year5	Country5
1996	USA	1997	USA	1998	USA	1999	USA	2000	USA
1996	USA	1997	USA	1998	USA	1999	USA	2000	USA
1996	USA	1997	USA	1998	USA	1999	USA	2000	USA
1996	USA	1997	USA	1998	USA	1999	USA	2000	USA
1996	USA	1997	USA	1998	USA	1999	USA	2000	USA

Librarians help shepherd and curate datasets.

Year	Country	Year2	Country2	Year3	Country3	Year4	Country4	Year5	Country5
1.996	1.520	1.100	1.112	2.009	1.626	0.623	4.524	0.811	1.033
2.964	3.991	1.911	1.909	1.999	1.999	1.420	1.039	1.688	1.688
0.963	0.882	0.282	0.795	0.908	0.467	0.146	0.327	0.172	0.411
0.176	0.142	0.034	0.160	0.132	0.170	0.040	0.148	0.043	0.144
1.369	1.271	0.475	1.709	1.111	1.931	0.862	1.962	0.908	0.919
12.208	17.318	18.910	0.903	0.909	1.378	38.989	47.449	12.921	11.818
0.441	0.319	0.144	0.440	0.299	0.279	0.247	0.246	0.236	0.246
1.298	1.284	1.384	1.980	1.908	1.262	0.642	1.124	1.430	0.797
1.089	0.644	1.167	2.114	2.902	2.119	1.703	1.910	1.869	1.869
0.314	0.188	0.026	0.139	0.122	0.036	0.088	0.270	0.099	0.099
0.633	1.792	1.282	1.267	0.912	1.039	4.067	0.483	2.201	0.904
0.368	0.075	0.009	0.166	0.207	0.196	0.002	0.241	0.061	0.122
0.247	0.112	0.009	0.414	0.210	0.211	0.077	0.208	0.044	0.101
0.244	0.142	0.041	0.270	0.296	0.176	0.047	0.123	0.054	0.104
0.328	0.180	0.036	0.267	0.160	0.160	0.003	0.229	0.030	0.119
0.214	0.090	0.002	0.167	0.200	0.090	0.002	0.171	0.064	0.104
0.802	0.495	0.196	0.836	0.468	0.618	0.206	0.916	0.238	0.365
0.462	0.128	0.000	0.426	0.528	0.194	0.007	0.973	0.132	0.201
1.444	1.044	0.903	1.407	1.427	0.823	0.891	1.161	0.964	0.980
1.428	2.101	1.101	0.861	0.860	0.847	0.267	2.288	0.811	0.811
0.792	0.261	0.070	0.260	0.201	0.110	0.107	0.177	0.063	0.071
1.248	1.026	0.962	0.923	1.702	1.701	0.003	1.167	0.894	0.794
0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.034	39.048	0.275	1.126	0.395	0.561	0.261	0.680	0.263	0.421
0.262	0.746	0.050	0.324	0.300	0.206	0.119	0.326	0.074	0.195
0.462	0.922	0.472	0.719	0.632	0.208	0.121	0.440	0.102	0.192

Librarians consult on data management plans.

Small science, single PI, small lab

P. Bryan Heidorn, while Program Director at NSF, demonstrated that small science = 80% * (awards under \$350,000)



By small science, we mean usually a single PI (or one who may collaborate with another PI) in a fairly small lab setting, where there is likely on average **2-6 graduate students/post docs** and *possibly* support from the department for a part-time lab manager or secretary.

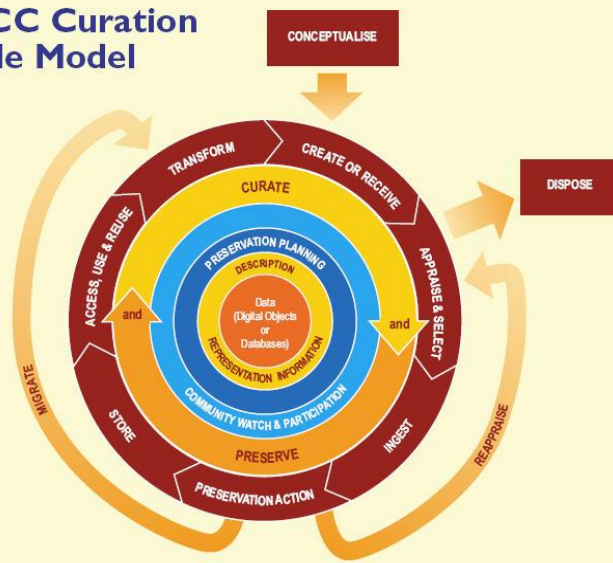
Small science can encounter unintentional consequences

Small science researchers self report: no specific person for data management/curation; **data is likely saved to hard drives in the lab** and backed up on CDs, usually by the students. While students have received “research integrity” training (which focuses on making data available upon request by funder, publisher, or FOIA, etc.) it is **not likely that anyone could retrieve usable data easily** or quickly.*

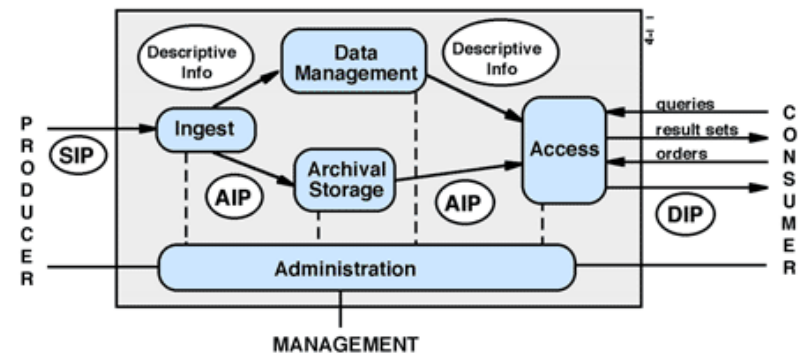


*D. Scott Brandt. Provost Fellowship, Final Report 2009 (*unpublished*)
all pictures Purdue University

The DCC Curation Lifecycle Model



Because single PIs tend not to have mature data curation in place, they are not likely to see where to enter data curation continuum



2006: Founded D2C2 to further investigations, organize research and leverage collaborations



Distributed Data Curation Center:

- Created “research arm” for Purdue Libraries
- Established recognizable mode for collaboration with research faculty on campus
- Focused research on data curation problems, distributed environments, **single PI/small lab**

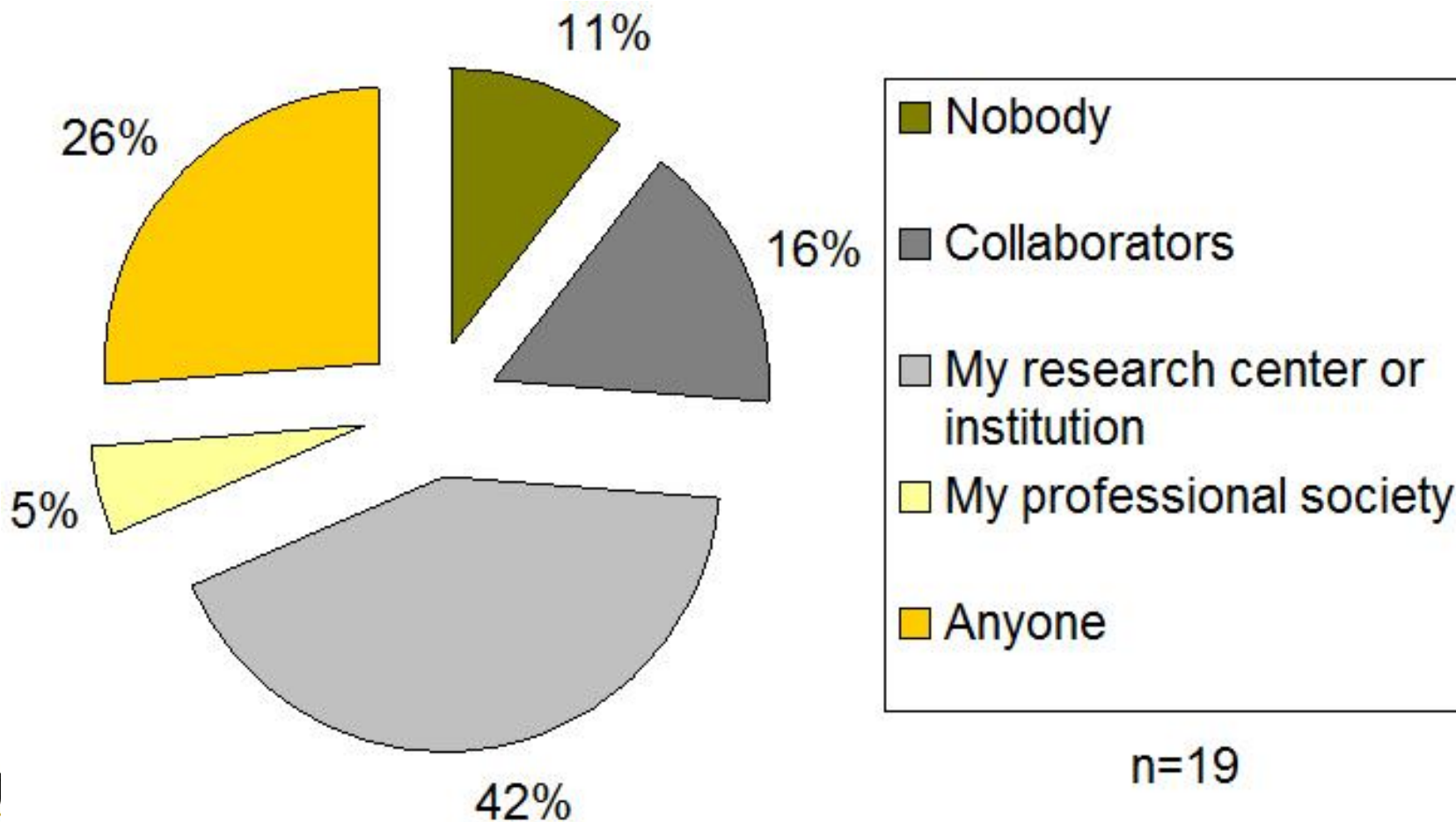
Producer





Project Title: "**Investigating Data Curation Profiles Across Multiple Research Disciplines.**" Investigators in the Distributed Data Curation Center in the Libraries at Purdue University, and the University of Illinois, Urbana-Champaign will address the question *“which researchers are willing to share data, when, with whom, and under what conditions?”* The team will produce case studies of researcher data/metadata workflow, curation profiles describing policies for archiving and making available research data, a matrix to compare parameters across disciplines, system requirements for managing data in a repository, and recommendations for implementing results under diverse systems. The project will describe the roles of librarians and identify the skill sets they need to facilitate scholarly communication and data sharing.

With whom would you share your data *after* the data have been analyzed?



Prioritize your needs for the following types of services

The ability for researchers within my discipline to easily find this dataset



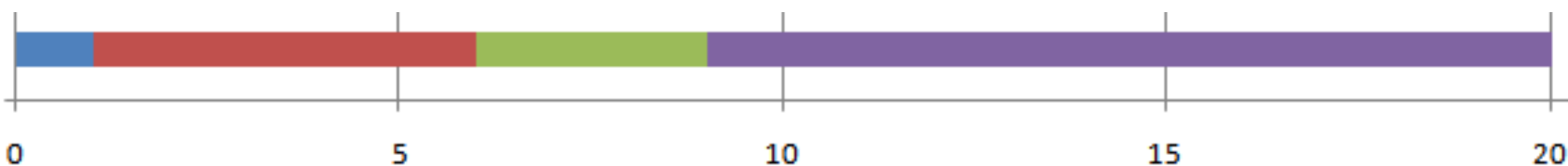
The ability for researchers outside of my discipline to easily find this dataset



The ability to cite this dataset in my publications



The ability for people to easily discover this dataset using Google



■ Not a priority

■ Low priority

■ Medium priority

■ High priority

■ I don't know or N/A

Who is willing to share what with whom?

What data sets do they have?

How are they managing these data sets?

What would they like to do with them?



Data Curation Profiles Toolkit



[About](#) |
 [News](#) |
 [Submit a Profile](#) |
 [Completed Profiles](#) |
 [Workshops](#) |
 [Forums](#) |
 [Resources](#)

Welcome to the Data Curation Profiles community!

Welcome to the community for Data Curation Profiles Toolkit!

This website is an environment where academic librarians of all kinds, special librarians at research facilities, archivists involved in the preservation of digital data, and those who support digital repositories can find help, support and camaraderie in exploring avenues to learn more about working with research data and the use of the Data Curation Profiles Tool.

A Data Curation Profile is essentially an outline of lifecycle within a research project. The Profile and environment of scholarly communication, especially further upstream than previously imagined. If research to data sets or collections, what does that mean for

Data Curation Profiles can:

- provide a guide for discussing data with researchers
- give insight into areas of attention in data management
- help assess information needs related to data
- give insight into differences between data in different disciplines
- help identify possible data services

LOGIN

Download the Toolkit

PURDUE
UNIVERSITY
LIBRARIES



DDCC
Distributed Data Curation Center

Download the Data Curation Profiles Toolkit

The Data Curation Profile Toolkit is available for anyone to download and use. However, we require that you register by creating an account on this site first. The Data Curation Profiles Toolkit is composed of four documents:

1. **The User Guide** - ([login to download](#))
2. **The Interviewer's Manual** - ([login to download](#))
3. **The Interview Worksheet** - ([login to download](#))

The User Guide provides information about the Data Curation Profiles, including background information, the purpose and use of Data Curation Profiles, and directions on how to construct a Data Curation Profile.

The Interviewer's Manual provides the framework for the interview. It contains text and questions to be read to the participating researcher over the course of the interview. Some of the questions to be asked will be in response to the answers given by the researcher in the Interview Worksheet (see below).

The Interview Worksheet is to be given to the researcher by the interviewer at the start of the interview. It is the worksheet that the participating researcher will fill out over the course of the interview. In addition to capturing important information, the responses provided by the researcher will serve as the basis for further discussion during the interview.

4. **The Template** - ([login to download](#))

The Data Curation Profile Template describes the structure of the Data Curation Profile. Each section or subsection within the Data Curation Profile template contains a brief definition of the information that is needed to populate an individual Data Curation Profile for the participating researcher.

Stage	Output	Typical File Size	Format
Pre-laboratory	Preliminary results of experiments		.xls
Record-Keeping	Description of experiments / data	NA	
Processed	Consolidated files for statistical analysis		
Analyzed	Statistical analysis Results of line statistical analysis		
Summarized	Graphs, charts, tables for use in articles / outputs		
Published	Note: The data specifically designated by the scientist to make publicly available are shaded in gray. Empty cells represent no response.		

Target data for sharing
The target data for sharing is the summarized data statistically significant effect on the cell's capacity to do research.

Use/re-use value of the data
The scientist notes that the data with the most value to other researchers. However, he notes that if an individual is interested in conducting some kind, the raw data might also be useful.

Contextual narrative
The data is considered to be static as it does not change or grow after digital data set is primarily composed of multiple tabular files in Excel not provide an exact size of the dataset but stated that the average from 20 to 50 KB. He was not sure of how many files comprise.

The data outputs for this dataset include Excel and SAS files. The scientist also maintains a lab notebook (in print) of analyzed statistics from the Excel and SAS files. If the data generated in the lab or by scientist

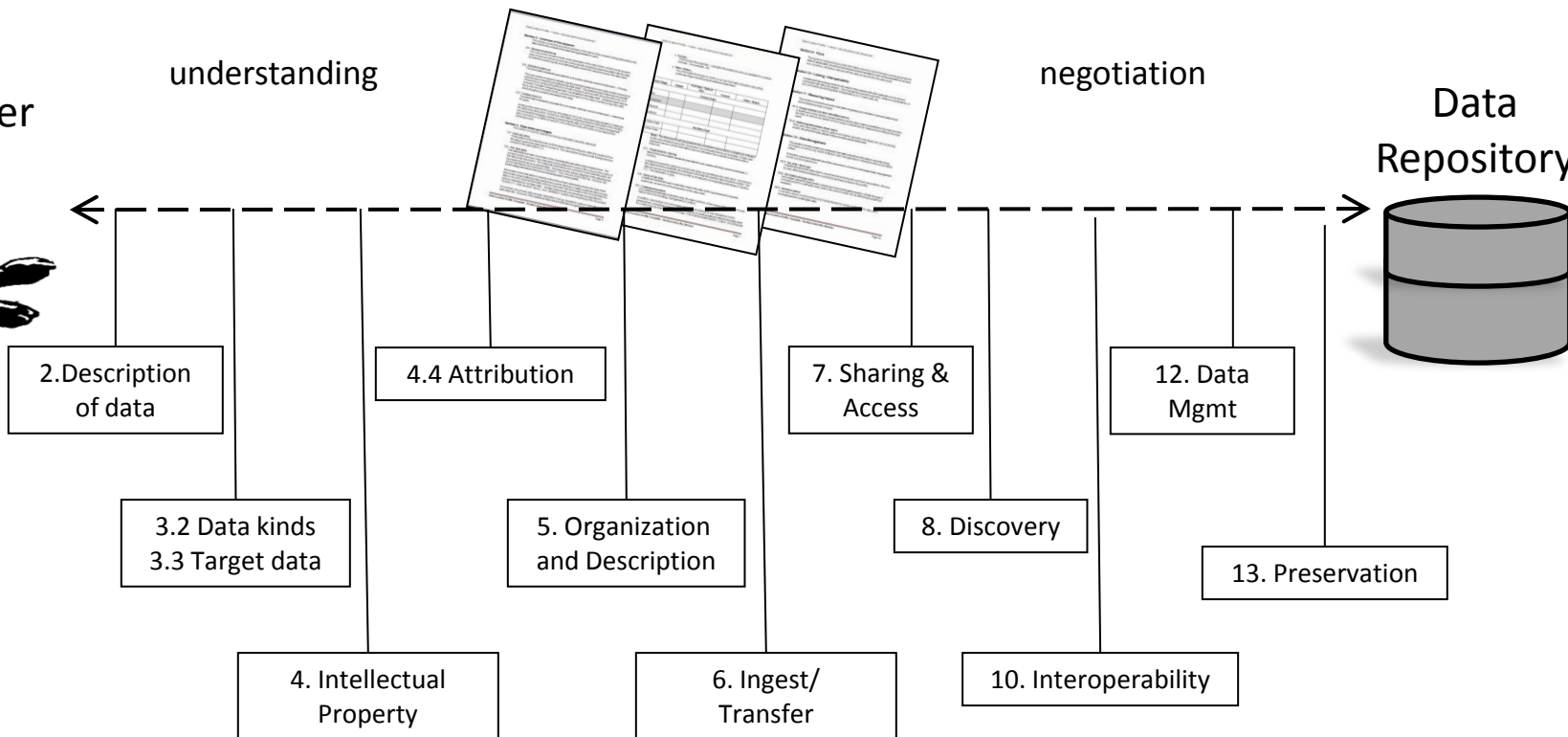


Researcher

understanding

negotiation

Data
Repository



Human Cell Defense Systems

The categories in the "data stage" column listed in the table below were developed by the authors of this data curation profile. The data specifically designated by the scientist to make publicly available are indicated by the rows shaded in gray.

Data Stage	Output	Typical File Size	Format	Other / Notes
Exploratory	Preliminary results of experiments		.xls	Data resulting from a series of small experiments; often stored in a spreadsheet.
Record-Keeping	Description of experiments / data	NA	Lab notebook - print	The experiment, along with the data, is described and captured in a lab notebook. This includes print-outs of results from scientific instruments.
Processed	Consolidated files for statistical analysis.		.xls	Data from experiments that show promise are reorganized into larger spreadsheets and prepared for analysis.
Analyzed	Statistical analysis		.xls, .sas	Depending on the nature of the statistical tests needed, the raw data is either analyzed in Excel, SAS, or both.
Summarized	Results of the statistical analysis	Between 20-50 kb	.xls	Data shown to be statistically significant.
Published	Graphs, charts and tables for inclusion in articles / output			The summarized data is usually organized into a graph and submitted with journal articles.

Note: The data specifically designated by the scientist to make publicly available are indicated by the rows shaded in gray. Empty cells represent cases in which information was not collected or the scientist could not provide a response.

Target data for sharing

The target data for sharing is the summarized data – the data that has been shown to have a statistically significant effect on the cell's capacity to defend against oxidative stress.

Use/re-use value of the data

The scientist notes that the data with the most value to other researchers are the summarized results. However, he notes that if an individual is interested in conducting a meta-analysis of some kind, the raw data might also be useful.

Contextual narrative

The data is considered to be static as it does not change or grow after it has been collected. The digital data set is primarily composed of multiple tabular files in Excel format. The scientist could not provide an exact size of the dataset but stated that the average size of his Excel files range from 20 to 50 KB. He was not sure of how many files comprise the dataset.

The data outputs for this dataset include Excel and SAS files as well as graphs incorporating analyzed statistics from the Excel and SAS files.

The scientist also maintains a lab notebook (in print) describing his experiments and containing print outs of the data generated in the lab or by scientific instruments. He also requires his

The Data Table:
Identifying data stages helps to clarify what might/can/should be disseminated, archived, preserved

Worksheet

2. How long would your data set be useful or have value for you or others if it were to be preserved?

- My dataset does not need to be preserved.
- Less than 3 years.
- 3 years or more but less than 5 years.
- 5 years or more but less than 10 years.
- 10 years or more but less than 20 years.
- 20 years or more but less than 50 years.
- 50 years or more but less than 100 years.
- Indefinitely.

Have the interviewee answer question 2 on the worksheet relating to the length of time the data should be preserved. Then ask the interviewee to talk a little about his/her response:

- Why did you select this length of time?

Manual

Template

Section 13 - Preservation

This section contains information about the needs / desires of the data client regarding the preservation of the data set under discussion.

If needed, a general statement about the researcher's preservation needs can be inserted here.

13.1 - Duration of preservation

A statement about the length of time the data is to be preserved. The duration may be event based rather than time based, though estimation for the length of time related to the event should be noted where possible.

U1N1S2 – (reviews the length of preservation question) – I think it is probably this.

I – Okay, 10 to 20 years.

U1N1S2 – Well I have data from my PhD that still would be useful if I could still access it. So that's from 1994, so yes, 10 to 20 years.

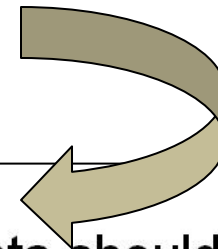
I- Okay. Is there a particular reason for the time frame as opposed to say 50 or 100 years?

U1N1S2 – I guess I'm hoping that we are able to do the modeling by then.

I – I see, so once the model has been worked out and you are very confident that the model does what you want it to do, then the underlying data used to get to that point would not be as valuable anymore?

U1N1S2 – I think so, yeah. And other things in that time period. actually used. So...

Transcript



13.1 - Duration of preservation

The scientist believes that her data should be preserved for 10 years or more but less than 20 years. The scientist believes that the mathematical model they are working on will be perfected within this timeframe. When the model is considered complete the data will lose much if not all of its value. In addition, the device used in generating the data may become obsolete within this time period and replaced, again diminishing the value of the data.

Profile

Data Curation Profiles Toolkit



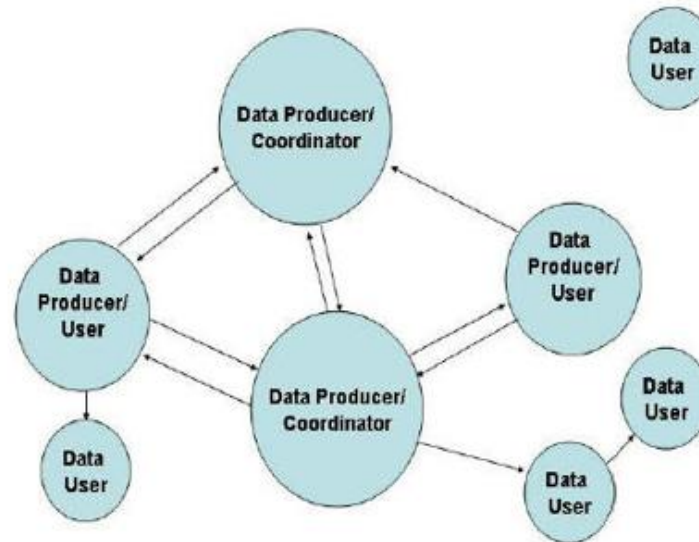
About ▾ News Submit a Profile Completed Profiles Workshops ▾ Forums Resources

Home

Module 10: Data Sharing

Intro Importance Data Sharing Table Interviewing Worksheet Sharing Review

This module presents information on data sharing and includes some helpful exercises for determining a researcher's conditions for sharing their data. Before beginning this module, it is recommended that you have an understanding of the *data lifecycle* as presented in module 9.



Source: http://www.urisa.org/publications/journal/articles/when_data_sharing

Profiles



DO NOT HANG ITEMS
FROM SPRINKLER HEAD
THIS WILL CAUSE
FLOODING

Data curation is an active process...

How this is manifested in PURR

<http://research.hub.purdue.edu>

The screenshot shows the Purdue University Research Repository (PURR) website. At the top, the Purdue University logo is on the left, and navigation links for 'Login', 'Register', and 'Report a bug' are on the right. Below the logo, the text 'Purdue University Research Repository | PURR' is displayed. A secondary navigation bar includes 'Home', 'Browse Content', 'Projects', 'Get Started', and 'Contact Us', along with a search bar. The main content area features a large banner with the title 'What Does My Data Management Plan Need to Address?' and a sub-headline 'What Does My Data Management Plan Need to Address?'. The banner includes a video player showing a person in a hard hat and safety vest working on a laptop, with a wind turbine in the background. Below the banner, there is a 'Learn More' button. A 'In the News' section highlights a news item: 'The National Science Foundation now requires all proposals to submit a Data Management Plan' with a 'Read More' link. Below this, there are two columns of content. The left column is titled 'Start Your Research Project' and contains three items: 'Read the DMP Requirements', 'Create a Group for Your Project', and 'Upload Research Data to Your Project'. The right column is titled 'Do you have a question?' and features a 'Ask a Librarian' form with fields for 'Enter your email address' and 'Your Question/Message', and a 'Send' button. At the bottom left, the 'PURDUE UNIVERSITY LIBRARIES' logo is visible.

PURDUE UNIVERSITY | Purdue University Research Repository | PURR

Home Browse Content Projects Get Started Contact Us Search

What Does My Data Management Plan Need to Address?

There are many things that need to be included in your data management plan. Watch our video tutorials, read the step by step instructions, or view a completed data management plan to help you get started. You can also click the link below to start now.

[Learn More](#)

In the News: The National Science Foundation now requires all proposals to submit a Data Management Plan [Read More >](#)

Start Your Research Project

- Read the DMP Requirements**
Read the detailed requirements for your data management plan. The requirements are very specific and should help to clear up any confusion regarding your own data management plan. [Get Started >](#)
- Create a Group for Your Project**
Groups are an easy way to share content and conversation, either privately or with the world. Create your new project group as a collaborative environment for sharing information and as a portal for your project team. [Get Started >](#)
- Upload Research Data to Your Project**
Become a contributor and share your work with your team and the community! Contributing content is easy. Our step-by-step forms will guide you through the process. [Get Started >](#)

Do you have a question?

Ask a Librarian

Send us your email question.

PURDUE UNIVERSITY LIBRARIES

PURR functionality & service

PLAN

Data mgmt
planning

DEVELOP

Creating
projects,
collaborating

ENHANCE

IF grant
awarded,
more space

PUBLISH

Data
submitted

for
publishing
/archiving

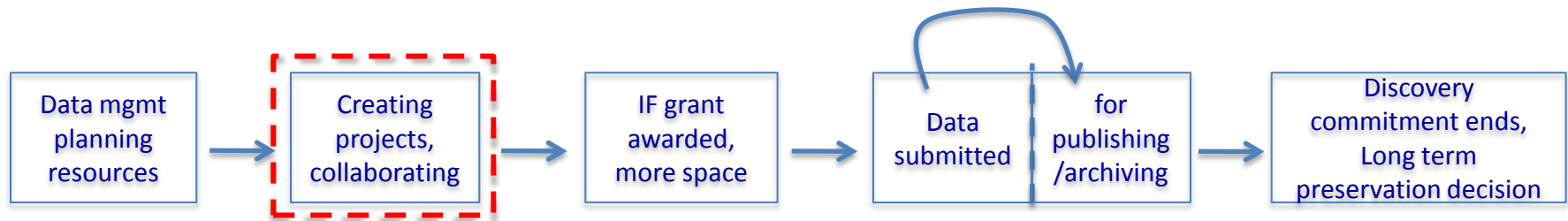
DISSEMINATE

Discovery
commitment ends,
Long term
preservation decision

Researchers use PURR for help with data mgmt plans via Pre-Awards, workshops and promotion, and by word-of-mouth

Librarians consult on data management plans.

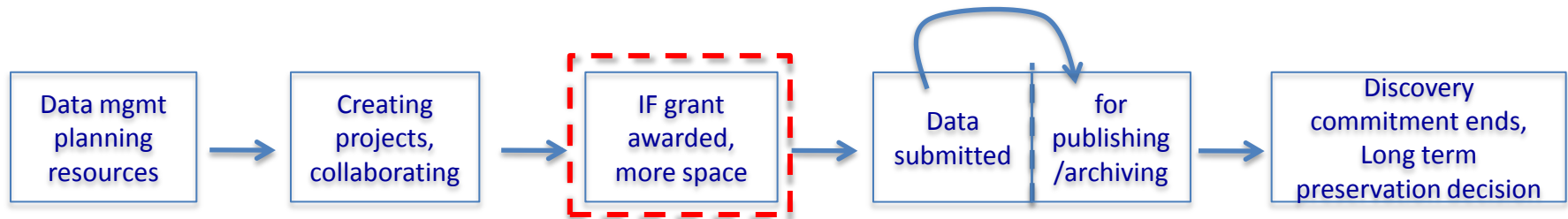
PURR functionality & service



Researchers can create projects at any time, invite others to join... the goal is to help facilitate research development

Librarians can contribute to or collaborate on projects.

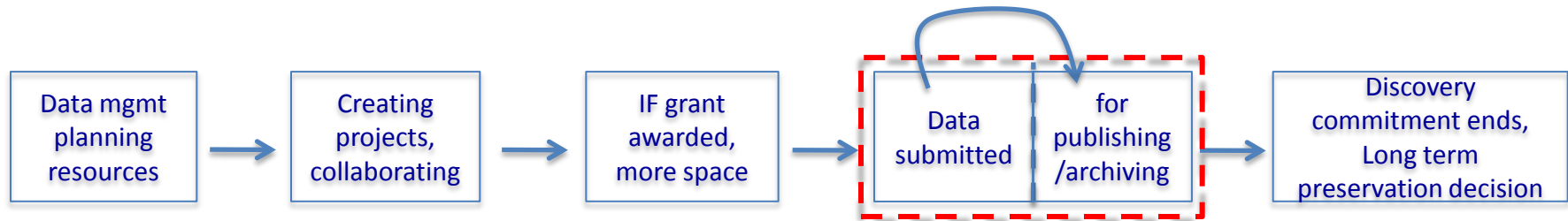
PURR functionality & service



Once a grant is awarded, researchers get an increase in space allocation and length of time for project and data

Librarians consult on data management.

PURR functionality & service



To make data sets publicly discoverable and available, there is a submission and “publishing” process

Librarians help shepherd and curate datasets.

Networks and Matrix Computations (nmcomp)

General project by David F Gleich

- Updates
- Info
- Team 2
- Files 5
- Publications**
- To-do
- Notes

Manager options:

- Edit project
- Invite people to join
- Learn about projects

To make data sets publicly discoverable and available, there is a submission and “publishing” process

Publications

This project has yet no publications. [Start a new publication](#)

How the publication process works...

- 1 Choose and arrange your content



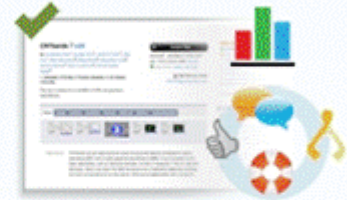
Select content from your project files. This may be a single file or multiple files bundled together. You may also add supporting documents e.g a user guide.

- 2 Describe publication and submit for review



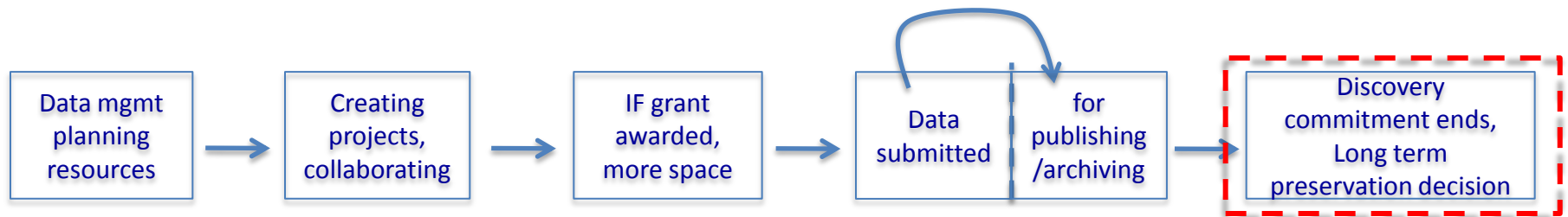
Next you compose your publication page, adding title, abstract, description, authors and other metadata. You may also add tags and screenshots.

- 3 Publish, archive, or save for review



When draft is ready, you may release your work publicly, archive it, or save draft for internal review. Public release comes with a digital object identifier and requires administrator approval.

PURR functionality & service



PURR policy allows for a specified time for discovery, and then decisions are made regarding long-term preservation

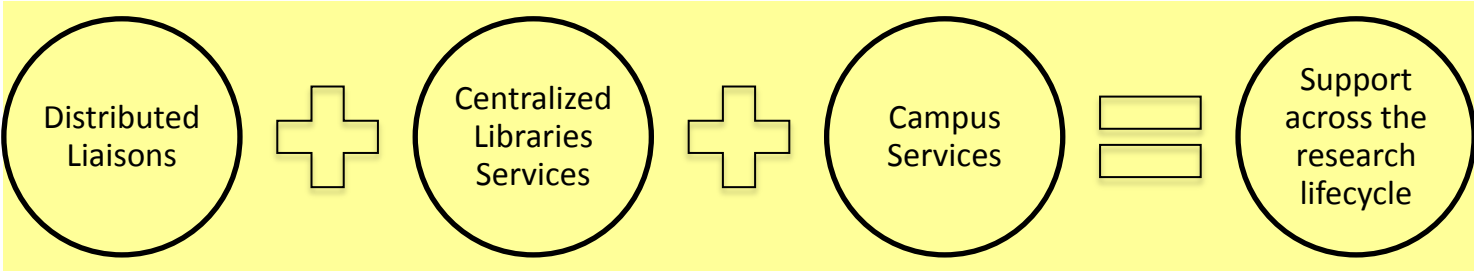
Librarians provide education on depositing and finding data.

Relevancy





PURDUE
UNIVERSITY
LIBRARIES



PURDUE
UNIVERSITY
LIBRARIES



PURDUE UNIVERSITY Office of the Vice President for Research

Home Integrity/Regulatory/QA Research Development Funding Partnerships Center

OVPR ▸ Research Development

Services & Resources

- Research Development Home
- Funding
- Limited Submissions
- Proposal Preparation

Research Development

The goal of the OVPR Research Development staff is to assist faculty in the proposals. **Our staff** provide a broad range of services and resources related some of the ways we can assist.

FUNDING
The funding page provides information on internal, ex

PURDUE UNIVERSITY Sponsored Program Services

General Info Pre-Award Post-Award Contract Management Data Training Links SPS He

Business Services Administrative Information Services (AIS)

Business@Purdue

Business@Purdue News

Sponsored Program Services

News

- FY2011 A-133 Report (**New**)
- Archived News
- Subrecipient Monitoring Procedures

PURDUE UNIVERSITY Sponsored Program Services Pre-Award Services

General Info Pre-Award Post-Award Contract Management Data Train

Pre-Award: General Info Compliance Proposals Sponsors

Proposal Preparation

- Interdisciplinary Proposals
- OVPR Proposal Preparation Resources
- OVPR Proposal Preparation Assistance



OVPR

Experts

3 Feb 2011 Data Management Plan Workshop

What about my current proposal?

- **Option 1: use the Purdue Data Management Plan**
 - Libraries/OVPR will provide a citable overview of basic, long term issues
 - You append details relevant to your project guided by questions provided by Librarians
 - Librarians are available to provide review (Nelson, M. Stowell Bracke, and J. Carlso)
- **Option 2: exploit data management in place for your discipline**
 - Example: data management resources at the Consortium for Political and Social Research
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>
- **Option 3: personalized solution based on your needs, provide and control**

What about my current proposal?

- **If you elect to use the Purdue University Research Repository (PURR)**
 - The quotable overview is here:
<http://vpr.hubzero.org/about/usehub>
 - Pre-award staff also have a short paragraph suitable for inclusion in the “Facilities, Equipment and Other Resources” section of your proposal
- It explains the current prospect of a centrally supported PURR resource available at no direct cost to the project.

IT

OVPR

Experts

7a-d are related to Export Control.

	Does this project require sponsor approval of publications?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
	b) Does the announcement restrict participation based on citizenship?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
	c) Does this proposal contain technical data within the statement of work that are restricted for purposes specific to military or space applications?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
	d) Does this proposal contain technical data within the statement of work that are confidential or proprietary to a company or Purdue (i.e. requires a confidentiality agreement)? If yes, list the relevant pages. [Redacted]	<input type="checkbox"/> Yes	<input type="checkbox"/> No
8	Are you receiving private <u>health information</u> that is protected by HIPAA?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Project Data			
1	Is there a <u>conflict of interest</u> ?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
2	Is this a Bowen Lab project?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
3	Is this a Kepner Facility project?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
4	List any approved University Centers affiliated with this proposal. [Redacted]		
5	Please identify all ITaP units which have or will contribute to this project: <input type="checkbox"/> Envision Center for Data Percentualization <input type="checkbox"/> Rosen Center for Advanced Computing <input type="checkbox"/> Other ITaP Units <input type="checkbox"/> None		
6	Is this proposal a resubmission of a previously submitted application?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
7	Are you an NSF Beginning Investigator?	<input type="checkbox"/> Yes	<input type="checkbox"/> No
8	Does the funding agency require a management plan for data or digital products produced using the grant? <input type="checkbox"/> Yes, and we plan to use the <u>Purdue University Research Repository (PURR)</u> https://research.hub.purdue.edu/ <input type="checkbox"/> Yes, and we plan to use another option <input type="checkbox"/> No, funding agency does not require		
9	Will historical sites be affected? If Yes, please provide an explanation:	<input type="checkbox"/> Yes	<input type="checkbox"/> No
10	If this project has an actual or potential impact on the environment, has an exemption been	<input type="checkbox"/> Yes	<input type="checkbox"/> No

ty



PURDUE UNIVERSITY Office of the Vice President for Research

Home Integrity/Regulatory/QA Research Development Funding Partnerships Center Support

OVPR > Main

Integrity Human Research Protection Program News

Conflict of Interest +

Research Integrity + Amendments to Exempt Studies



PURDUE UNIVERSITY **HIPAA Compliance**

Notice of Privacy Practices Covered Components Guidelines Forms and Procedures Training Patient Rights

PURDUE UNIVERSITY Office of the Registrar

Students Commencement Residency Scheduling Veteran Affairs myPurduePlan

Information links

Active PWL Programs
Advisors
Calendars
Contacts

FERPA and Residency Contact

Family Educational Rights and Privacy Act (FERPA)
Residence Classification, Information and Changes



PURDUE UNIVERSITY Office of the Vice President for Research

Home Integrity/Regulatory/QA Research Development Funding Partnerships Center Support

OVPR > Research Integrity/ Regulatory Affairs/ Quality Assurance > Research Oversight

Integrity Export Control Regulations

Conflict of Interest +



OVPR

Experts

Purdue University Research Repository | PURR

Will you be uploading any data that may be considered by the University to be sensitive or restricted? *REQUIRED*

Yes, this project may involve uploading sensitive or restricted data to PURR.

Please indicate the type of data that may be involved. Check the box if you're not sure.

This project may involve government-restricted, export-controlled, or proprietary company information (without permission).

This project may involve data that is governed by an Institutional Review Board (IRB) Approval Protocol.

This project may involve HIPAA data or Protected Health Information.

This project may involve FERPA data or student records.

IT

Faculty



PURDUE UNIVERSITY Office of the University CIO

About Academic Administrative ITaP Governance Org Charts CIO Update Datacenters

Login

Office of the University Chief Information Officer



Dr. Gerry McCartney
Vice President for Information Technology & CIO
and Oesterle Professor of Information Technology

The Office of the University Chief Information Officer strives to develop and deliver world-class computing and digital resources to Purdue faculty, staff, and students. The goals of the office are to enable research in science and engineering, provide classroom technologies to support and improve learning, and present innovative tools for engagement.

PURDUE UNIVERSITY Information Technology at Purdue
Rosen Center for Advanced Computing

Home News & Events Projects User Info Computation Storage Purchase [view more](#)

XSEDE

Extreme Science and Engineering
Discovery Environment

Purdue is a partner in XSEDE, the most advanced, powerful and robust collection of integrated digital resources and services



HUB PLATFORM FOR SCIENTIFIC COLLABORATION
CREATED BY PURDUE UNIVERSITY

HOME GET STARTED DOCUMENTATION ABOUT

What is HUBzero?

HUBzero is a platform used to create dynamic web sites for scientific research and educational activities. With HUBzero, you can easily publish

Start your own HUB

Use HUBzero to create your own site. [Download](#) our open source release or have a hub setup and hosted for you via [Purdue's hosting service](#).

HUBBUB 2012
the HUBzero conference

HUBbub 2012 will be held at the Unive

HOW HUBS WORK

HUBzero is a new way for scientists and engineers to publish and share information. The latest hub will be used to study the causes and effects of earthquakes for the Network for Earthquake Engineering Simulation, or NEES. This hub joins others focused on topics such as nanotechnology, pharmaceutical products, advanced manufacturing, cancer care, assistive technologies for people with disabilities, heat-transfer issues in engineering, and several others. New hubs are being created at a rate of about one per month.

1) A scientist named Hugh creates software for scientific modeling and uploads it to the hub to share with colleagues, much as he would share a video on YouTube. Other scientists and engineers can use the tool and rate or comment on it.

2) Another researcher, Sue, runs her data on the new software tool using a simple Web interface. The demanding computations are done using cloud computing, and are automatically farmed out to available computers across the nation. Sue's data is sent to supercomputers connected to the TeraGrid, Open Science Grid, or the new DiaGrid, and she receives her results within minutes.

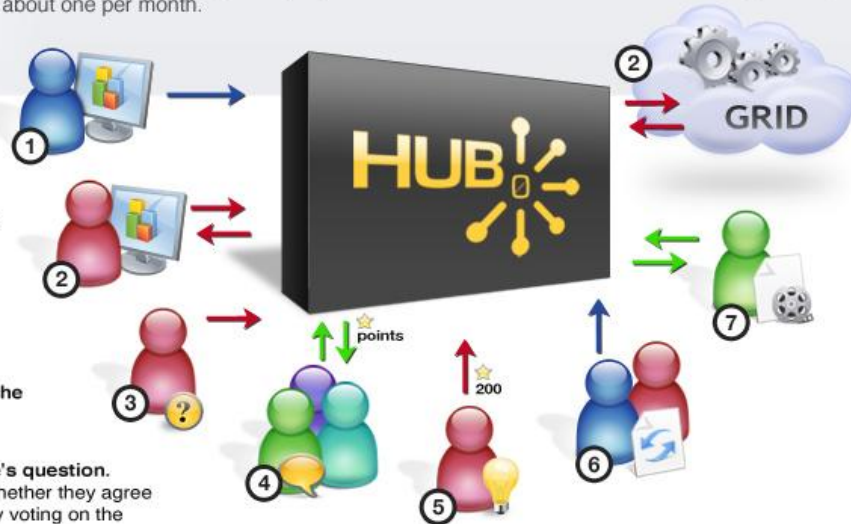
3) Sue has questions about her results, so she asks the community of scientists on the hub what they think about the finding.

4) A third researcher, Drew, posts a possible answer to Sue's question. Then, other scientists vote "thumbs up" or "thumbs down" on whether they agree with Drew's answer. Drew and the other scientists participate by voting on the best answer, all receiving hub points for their efforts.

5) Sue also has an idea for a new feature for the software tool, so she adds her idea to a wish list for the software on the hub. Because this is a feature she really wants, she offers a bounty of 200 of her hub points for the feature.

6) Sue and Hugh decide to collaborate on their next project, which is successful, so they write a research paper and give a presentation at a conference. They post the presentation slides on the hub and then track how many users view the materials.





7) Drew presents the presentation and uses some of the slides in his class to explain this new area of research to his students. He creates a homework assignment which his students use Hugh's simulation tool to investigate the phenomena, and then Drew posts the assignment on the hub for other educators



OVPR

Experts

Storage space for PURR projects currently covered by central resources

<i>Option:</i>	<i>For what:</i>	<i>Space Available:</i>	<i>For how long:</i>	<i>For who:</i>
 Default/trial projects	Just trying things out, or don't need much space	500 MB	3 years	Anyone with a Purdue Career Account
 Default data publications	Great for small publications	50 MB	10 years	Anyone with a Purdue Career Account
 Supported projects	Funded projects with PIs from Purdue	100 GB	Life of project	Purdue faculty with a verifiable grant or account number
 Supported data publications	Publishing work done on a funded project	10 GB	10 years	Purdue faculty with a verifiable grant or account number

Estimated cost of additional space *

<i>Option:</i>	<i>Space Available:</i>	<i>For how long:</i>	<i>For who:</i>	<i>Cost:</i>
Extra project space	As needed	Life of project	Anyone with a Purdue Career Account	\$2,102.57 per TB per year **
Extra publication space	As needed	10 years	Anyone with a Purdue Career Account	\$14,297.48 per TB

IT

Faculty



OVPR



Experts

Applying library science to information problems

“You know thirty years ago, good laboratory practice was... you took good notes, you took fifteen or twenty data points... and you had a nice little lab book. But we’ve scaled now to getting this mega amount of information and we haven’t scaled our laboratory management practices.”

“In an ideal situation we would somehow have some sort of standard under which we named things and stored things and kept track of things and we would, you know, have a way to get this information to our students.”



IT



Faculty

Identify



Utilize Data Curation Profiles to collect information about current data gathering, workflow and documentation

Assess



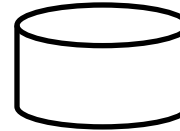
Review applicable data & metadata standards, as well as best practice to document workflow (e.g., USDA, DfataONE, AgMES, AGROVOC)

Analyze



Perform requirements analysis to determine specifications for new user demands which include enhanced data management & sharing

Model



Develop a data model that accommodates collection, recording and workflow of data as well as application of thesaurus/ontology

Demonstrate



Demonstrate through a use case that data can be ingested and used by a modeler; and compliant with DataONE API

Experts

IT

Problem: Water quality research is conducted by several groups which results in various amounts and sets of data which need to be organized

Faculty

Curiosity + Collaboration = Choreographic Expression

A History of Concert Dance at Purdue University 1931-2011

- The Queen is Dead -



Excerpts from the 2009 Purdue Contemporary Dance Company performance of "The Queen is Dead".

Choreographer's notes: "The Queen is Dead was created in about 3 weeks-- a very fast process for me. I started from a very different premise than where I ended. I used some concepts for the original movement vocabulary creation from the early ideas, and then used a different concept when actually constructing the piece. I used images and text from a collage on a table in a coffee shop where I was sitting to prepare for rehearsals, plus ideas about falling and physical limitations, and then it all came together when a mentor suggested using the actual music used for coronations in England. It was created for a different cast, in Milwaukee, Wisconsin, and then set on the PCDC cast. The original work had 2 men and 1 woman, but I think it totally worked with the 2 women and 1 man."

Date:	2009	Performers:	Mandy Hampton Carrie Meyer Paul Rutz
Venue:	Hansen Theatre, Yue-Kong Pao Hall, Purdue University	Musicians:	N/A
Choreographer:	Holly Jaycox (Division of Dance faculty)	Lighting Designer(s):	Krystle Smith
Composer:	N/A	Costume Designer(s):	Holly Jaycox
Sound Designer(s):	Seth Warren-Crow	Set Designer(s):	N/A
Source Music:	William Boyce Propellerheads	Videographer:	Kathy Evans (Visual Arts Librarian)
Other Collaborator(s):			

Subject: Faculty choreography, Dance performance

Type of content: Dance performance

Content format: H.264/MPEG-4

Content editor: Eugenia Kim

Content source: Division of Dance video archives

Publisher: Virginia Kelly Karnes Archives and Special Collections Research Center, Purdue University Libraries

Rights Policy:

This video is for recreational viewing only and is not meant to be reproduced nor used for research purposes. Please contact the [Virginia Kelly Karnes Archives and Special Collections Research Center](#) for inquiries about the original media.

ChoreoSave: A Digital Dance Preservation System Prototype

Eugenia S Kim
Purdue University Libraries, Purdue University
West Lafayette, IN, USA
eugeniakim@purdue.edu

ABSTRACT

This paper presents the design and implementation of a short-term digital dance preservation solution called *ChoreoSave*. The author created a prototype system that identifies what components comprise a dance work and how such components can be represented using EPrints software hosted on an Amazon EC2 instance. Suggestions for future development are discussed based on evaluation and challenges encountered in the course of research.

such as music through documentary footage and contextual information. This approach challenges traditional notions of capturing movement through the use of information organization principles. By forming a standard submission template with customization options for users, the appraisal, selection, and ingestion processes will become the users' responsibility, thereby streamlining the flow of content from creator to repository. The short-term characteristic coupled with a peer-based model facilitates a more

Problem: How are digital objects of dance identified, accessioned, and prepared for distribution? And what are the best practices and policies required to produce a multimedia-based dance history website?

Faculty

Roles



General Library Services

Reference: assessing complex information needs

- Meet researchers in the lab, office, classroom, Starbucks
- Discuss issues/problems of all aspects of scholarly communication
- Identify resources and ways to meet those needs

Instruction: enhance finding, evaluating, using research

- Explore activities related to creating and using information
- Partner with faculty to teach various illiteracies (info, data, etc.)

Collection mgmt: appraising local collections, including data

- Identify content for different dissemination modes
- Determine selection practices for new collections

Liaison: engaging researchers in new ways, in their environments

- Discuss research initiatives, projects, outputs...

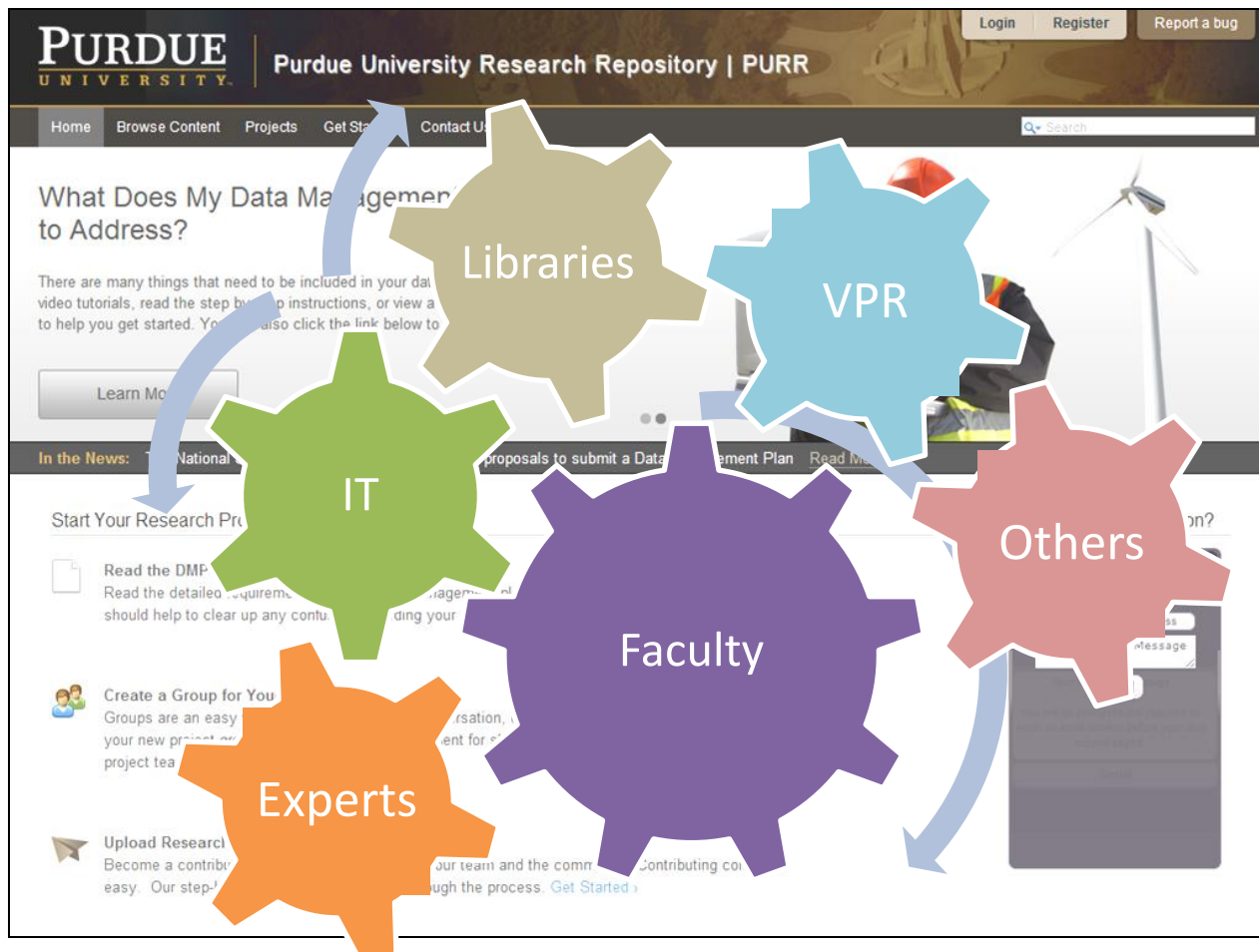
– Collaborate on funded research projects— apply library science

Specific Data Services

- Data reference
- Data mgmt planning
- Data consultation (may lead to collaborations/grants)
- Using PURR
- Promoting data DOIs
- Data mgmt education and information literacy
- Finding and using data
- Developing tools (DCP 2.0, DataBib, DMP-SAQ)
- Data visualization/GIS
- Developing data resources (LibGuides, tutorials)
- Linking data to articles and dissertations
- Promoting open access (Authors rights, IR deposit)*
- Leveraging publishing opportunities*
- Developing local collections*
- Collection mgmt of “e” (journals, data, archives)*
- Integrating systems * (i.e., finding data in Primo)

* As relates to data

Data curation is a process based on partnerships & collaborations





Questions?

