2015 DLF E-RESEARCH NETWORK

Webinar 1: RDMS At Your Institution

May 13, 2015
TODAY’S WEBINAR

• Guest Speaker
• Q & A
• Group activity slides
• Discussion
• Housekeeping
• Tweet about it! #eresearchnetwork @CLIRDLF
D. Scott Brandt, Professor of Library Science and former Associate Dean for Research in the Purdue University Libraries
“Digital Curation is defined as the **active** management and **enhancement** of digital information assets for **current and future** use.”

- active as opposed to passive, and that’s why we build data management services to consult and collaborate
- enhancement includes metadata, documentation and software to find & use, which takes time and effort
- discoverable by and available to wider audiences over longer periods of time, which is what libraries do…

National Research Council, Board on Research Data and Information
http://sites.nationalacademies.org/PGA/brdi/PGA_069853
We tend to focus on the “small science” researchers who report that they have no money or time for data management, and no specific person for data curation, other than graduate students who have received “research integrity” training, but it is not likely that anyone could retrieve usable data easily or quickly.
GUIDING PRINCIPLE

Multiple approaches needed:

• Our research showed that “Data curation services will need to accommodate a wide range of sub-disciplinary data characteristics and sharing practices.”

• Thus, Purdue combines a full range service institutional data repository and specialists to provide a comprehensive approach to data consultation and collaboration.

DATA CURATION PROFILES

Researcher

2. Description of data
3.2 Data kinds
3.3 Target data

4. Intellectual Property

4.4 Attribution

5. Organization and Description

6. Ingest/Transfer

7. Sharing & Access

8. Discovery

10. Interoperability

12. Data Mgmt

13. Preservation

Data Repository
ORGANIZATIONAL APPROACH

Distributed Liaisons

Centralized Libraries Services

Campus Services

Support across the research lifecycle

Faculty Liaisons
- Divisions: PSET, HLS, HSSEB

Data Management
- Data and Metadata Services
- Data Education Working Group
- Distributed Data Curation Center

OVPR
- Research Admin
- Sponsored Programs
- RCR & IRB

Disciplinary Faculty
- Graduate Students
- Undergraduate Students

PURR
- Digitization Center
- Archives & Spec Collections

ITAP
- Central IT
- HUBzero
- Rosen Center for Advanced Computing

Publishing
- University Press
- e-Pubs IR

Rights Management
- University Copyright Office

PURDUE UNIVERSITY LIBRARIES
Research Data

Specialists

Liaisons

Additional specialists

Faculty, students, staff

ITaP, OVPR

D2C2/PURR

BOOTS ON THE GROUND (INPUTS)
Creating projects, collaborating

IF grant awarded, more space

Data submitted for publishing/archiving

Discovery commitment ends, Long term preservation decision

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
### Questions

7. a) Does this project require sponsor approval of publications?
   - Yes
   - No

   b) Does the announcement restrict participation based on citizenship?
   - Yes
   - No

   c) Does this proposal contain technical data within the statement of work that purposes specific to military or space applications?
   - Yes
   - No

   d) Does this proposal contain technical data within the statement of work that is proprietary to a company or Purdue (i.e. requires a confidentiality agreement) relevant pages.
   - Yes
   - No

8. Are you receiving private health information that is protected by HIPAA?
   - Yes
   - No

### Project Data

1. Is there a conflict of interest?
   - Yes
   - No

2. Is this a Bowen Lab project?
   - Yes
   - No

3. Is this a Keppner Facility project?
   - Yes
   - No

4. List any approved University Centers affiliated with this proposal.

5. Please identify all ITaP units which have or will contribute to this project:
   - Envision Center for Data Perceptualization
   - Rosen Center for Advanced Computing
   - Other ITaP Units
   - None

6. Is this proposal a resubmission of a previously submitted application?
   - Yes
   - No

7. Are you an NSF Beginning Investigator?
   - Yes
   - No

8. **Does the funding agency require a management plan for data or digital products produced using the grant?**
   - Yes, and we plan to use the Purdue University Research Repository (PURR) [https://research.hub.purdue.edu/](https://research.hub.purdue.edu/)
   - Yes, and we plan to use another option
   - No, funding agency does not require

9. Will historical sites be affected?
   - Yes
   - No
   
   If Yes, please provide an explanation:

10. If this project has an actual or potential impact on the environment, has an exemption been
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.
To make data sets publicly discoverable and accessible, there is a submission and “publishing” process that requires metadata and documentation.
Most researchers agree that open access to data is the scientific ideal, so what is stopping it happening? **Bryn Nelson** investigates why many researchers choose not to share.
Welcome to Research Data

We apply the principles of library and archival sciences to meet the challenges and needs of researchers and students in finding, using, preparing, managing, curating and preserving research data. Working with our liaison librarians, we can help you plan, describe, disseminate, steward and archive your datasets.

Have a question? Contact us at: researchdata@lib.purdue.edu

OUR SERVICES

- Purdue University Research Repository
- Data Management Planning
- Metadata & Data Documentation
- Data Management & Curation Education
- Big Data @ Purdue Libraries
- Data Consultation
The Research Data Specialists provide leadership, coordination, and expertise in areas of data management and data education. We do this by:

- exploring, planning and designing applications, methodologies, tools and techniques to improve data management planning
- working closely with liaison librarians to promote and deliver data services, such as using Data Curation Profiles to assess data workflow
- coordinating data-related outreach efforts, such as using PURR for data publishing, with subject liaisons
- developing courses and instruction related to data information literacy
It is USAID’s expectation that data collected for this project will be made accessible... how can the Libraries help?
1. Data is collected in the field and passed on data steward
2. Data can be aggregated in PURR
3. Documentation is collected
4. Data indicators uploaded to the FTFMS database
5. Over time, other datasets are deposited with documentation in repositories such as DDL or PURR
IPNI ACCESS & REUSE PROBLEM

The Ag data landscape...

data created here
cannot be found or easily used here

No data mgmt plan

Methods and protocols not documented for sharing

No use of file naming conventions to support organization

Lack of consistent data standards

Workflows not well organized or documented

Need to specify target data for publication

Does not integrate open access tools for interoperability

Unclear how to attribute/cite

Not well described for wide scale use and discovery

Lack of access and use policies

No thought given to long term preservation

No thought given to long term preservation
Ag data curation pipeline...

A set of practices, tools and services that ensure the use and reuse of data over time

- **Plan & Design**
  Policies, practices and roles are defined with the intent of making data available to others.

- **Collect & Capture**
  Methods are well documented and community standards applied.

- **Interpret & Analyze**
  Open source software and tools are used to facilitate use.

- **Release & Publish**
  Policies and activities of dissemination follow most open access practices.

- **Discover & Use**
  Documentation and metadata ensure that data sets can be found and used by many.

- **Manage & Preserve**
  Knowledge is managed for long term preservation and retrieval.
Data Curation Pipeline for Sustainable Intensification Innovation Labs (DCP4SIIL): The Pipeline will comprise a framework for collecting data and associated metadata, a platform to provide easy data entry and capture, as well as analysis and dissemination, and guidelines for “enhanced” Good Lab Practice to ensure community development and adoption of standards and best practices in data management and curation.
PRODUCERS and CONSUMERS

Faculty, Graduate and Undergraduate Students
THINGS TAKE TIME

2003-4

Dean

Associates Dean

Library Faculty -40-

Directors of Office of Copyright, Planning, Information Technology

2006-7

Dean

Directors of Office of Copyright, Finance, and the University Press

2013-15

Dean

Directors of Office of Copyright, Finance, and the University Press

Assoc Dean for Research & Assessment

Data Specialists

D2C2/PURR

Assoc Dean for Academic Affairs

Assoc for Academic Affairs

Assoc Dean for Planning & Administration

Assoc Dean for Digital Programs and Information Access

Library Faculty -34-

Library Faculty -42-
Diffusion of innovations is a theory that seeks to explain how, why, and at what rate new ideas and technology spread through cultures. (Wikipedia)
OUTCOMES

Curiosity + Collaboration = Choreographic Expression
A History of Concert Dance at Purdue University 1931-2011

- The Queen is Dead -

Excerpts from the 2008 Purdue Contemporary Dance Company performance of "The Queen is Dead".

Choreographer’s note: "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 from the early ideas, and then used a different concept when actually constructing the piece. I was sitting in my car on a table in a coffee shop where I was sitting to prepare for rehearsals, plus ideas about the narrative structure and then it all came together when a mentor suggested using the actual music used for the ballet "The Three Billy Goats Gruff" created for a different cast, in Milwaukee, Wisconsin, and then get on the PCDC cast of women, but I think it totally worked with the 2 women and 1 man."

<table>
<thead>
<tr>
<th>Date:</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venue:</td>
<td>Hansen Theatre, Yue-Kong Pao Hall, Purdue University</td>
</tr>
<tr>
<td>Performers:</td>
<td></td>
</tr>
<tr>
<td>Musicians:</td>
<td></td>
</tr>
<tr>
<td>Choreographer:</td>
<td>Holly Jaycox (Division of Dance faculty)</td>
</tr>
<tr>
<td>Lighting Designer(s):</td>
<td></td>
</tr>
<tr>
<td>Composer:</td>
<td>N/A</td>
</tr>
<tr>
<td>Costume Designer(s):</td>
<td></td>
</tr>
<tr>
<td>Sound Designer(s):</td>
<td>Seth Warren-Crow</td>
</tr>
<tr>
<td>Source Music:</td>
<td>William Boyce Preludes and Fugues</td>
</tr>
<tr>
<td>Videographer:</td>
<td></td>
</tr>
<tr>
<td>Other Collaborator(s):</td>
<td></td>
</tr>
</tbody>
</table>

Phosphorus and Potassium Influence on Alfalfa Nutrition

This is a master file of merged excel files with data from several studies on the influence of phosphorus and potassium nutrition on alfalfa. It includes numeric data such as yield, plant mass, plant counts, and tissue concentration of various nutrients. The data shows that plants with low phosphorus and potassium are often limited in yield and that plants that are poorly fertilized and do not have adequate phosphorus and potassium have lower yields. Therefore, it is important to use the right amount of fertilizers for better production.

Location/Latitude and Longitude: Throckmorton Purdue Agricultural Center located 15 km south of West Lafayette, Indiana.

Researchers should cite this work as follows:
Jeffrey J. Volenec, (2012), "Phosphorus and Potassium Influence on Alfalfa Nutrition" (DOI: 10.4231/D3S60F79Z)
More Outcomes

1334 toolkit users
360 librarians trained

Used as class assignments/projects in library and information science schools

Profiles used to develop focus group discussions with faculty

Profiles as basis for cases studies of UMMS/WPI Curriculum Frameworks

Model for collection, validation and publication of a range of data-centric “capability profiles”:
Community Capability Model Framework
Working Group for the Research Data Alliance

Other universities use it to consult with researchers

Data Profiling for a Water Quality Lab (profiling multiple GAs data workflow)
2013 Association of College and Research Libraries’ Science and Technology Section Innovation Award: Distributed Data Curation Center and the Data Curation Profiles Toolkit and 2015 Association of College and Research Libraries’ Excellence in Libraries Award


Carlson, J. (2012) Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about their Data


Thank you!

Questions?

D. Scott Brandt

techman@purdue.edu
Research Data Management Planning & Design at Caltech

Dona Wrublewski and Gail Clement
Caltech Library System
Current State

- Data management has a low-visibility footprint here
  - Could be because all is well and Library is not (perceived to be) needed
    - Activities are decentralized and a lot is happening in silos OR
    - Activities are centralized but Library overlooked as a partner
  - Could be because needs are not yet established: no one asking
  - Could be because no one has faced compliance issues yet
Guiding principles for our work

• Don’t presume there is a top down interest in RDM at campus scale
  • Expected culture: Likely things are being handled in distributed fashion, by diverse research groups, in accordance with funder/publisher requirements

• Don’t presume Library’s role – need to ask key stakeholders if/how we can help

• All proposed new resources and services need to be based on evidence (strong culture of data driven decision making)

• All proposed new resources and services need to be funded (not taken out of existing, highly constrained Library budget)
Beginning the Process

- Assess awareness and degree of perceived need based on campus culture and practice
  - Where to start conversations (via structured interviews)
    - VPR
    - Provost
    - CIO
    - Faculty Board (Senate)
    - Grad Student Association
    - Post Doc Association
  - Find out if a Library role is welcome
  - If Library role welcome, how can we help?
  - Establish that new services will require additional funding
    - Already there are many competing demands on Library, coupled with static budget
Figure 1 The process of developing research data management services
Repository inspiration

Realistic Peers
- University of Wisconsin - Milwaukee
- James Madison University

Aspirational Peers
- Libraries Digital Conservancy
- Research Data Services, University of Colorado Boulder
- OSU, Oregon State University
HiPerGator. UF Research Computing; Early IBM Computer in Statistics Lab & Webb holds a roll of data. UF University Archives, UF Digital Collections
UF: CONTEXT

- Public, land-grant; one of 17 in the AAU
- One of the largest, most complex, comprehensive universities in the world
-“Data”: basic and applied sciences, humanities, arts, law, design, education, journalism, & more; 16 colleges; 150+ centers/institutes; 50,000+ students; 4,000+ faculty
- Public and translational research, broader impacts, funder requirements
- Historically underfunded
  - Best at being poor
  - Tactical, opportunistic, strategic → parallel activities
- Infrastructures are always local
  - Culture and rich tradition of collaboration & innovation
  - Focus on building together to enable radical collaboration
Context, principles, inputs, outputs-services, outcomes

• Superb technical infrastructure: HiPerGator, UF Digital Collections (powered by SobekCM), UFApps, GatorBox

• Excellent, evolving socio-technical infrastructure: people, policies, technologies, communities (building communities of practice)

• Libraries, Research Computing, Office of Research: collaboration to determine, implement campus-wide supports

• Data policy campus-wide (for implementation as best practices now, as a policy pre-2020)
You Are Here

Pilot Phase

Senior Campus Administrators’ Consensus

Figure 1: The process of developing research data management services
Menus: The Art of RDMS Development

Inspired by our Special Collections *Menus: The art of dining*
http://digital.library.unlv.edu/collections/menus/about-collection-0

- **Main Dishes (market price)**
  - Case Studies
  - (DAF) Data Asset Framework
  - (CARDIO) Collaborative Assessment of Research Data Infrastructure and Objectives
  - (DCP) Data Curation Profiles
  - Stakeholder Profiles, Personas, and Usage Scenarios
  - Development Workshops

- **Blue Plate Specials**
  - Embed Questions in Campus-wide Online Survey
  - Environmental Scan
  - Read Data Curation Profiles (on Purdue site)

  *Warning – half-baked approaches may be harmful to institutional health.*
THANK YOU!

- Webinar 2: June 10\textsuperscript{th}, 1:00-2:30 PM EST - “Data Management Needs”
- DLF Forum (Oct. 26-28)
  - Collaborate on a panel / presentation
  - CFP closes June 22
- CLIR Connect