9 September 2010

Extending our Reach: Librarians as Research Partners in the Emerging Digital Data Framework

Jake Carlson, Data Research Scientist, Purdue University Libraries

Extending our Reach: Librarians as Research Partners in the Emerging Digital Data Framework

Jake Carlson
Data Research Scientist
Purdue University Libraries & D2C2

Transportation Librarians Roundtable

Content

1. Data Curation and the Role of Libraries.
2. The Work of the Purdue Libraries and D2C2
3. My Work as a Data Research Scientist
4. What's coming next?

The Problem

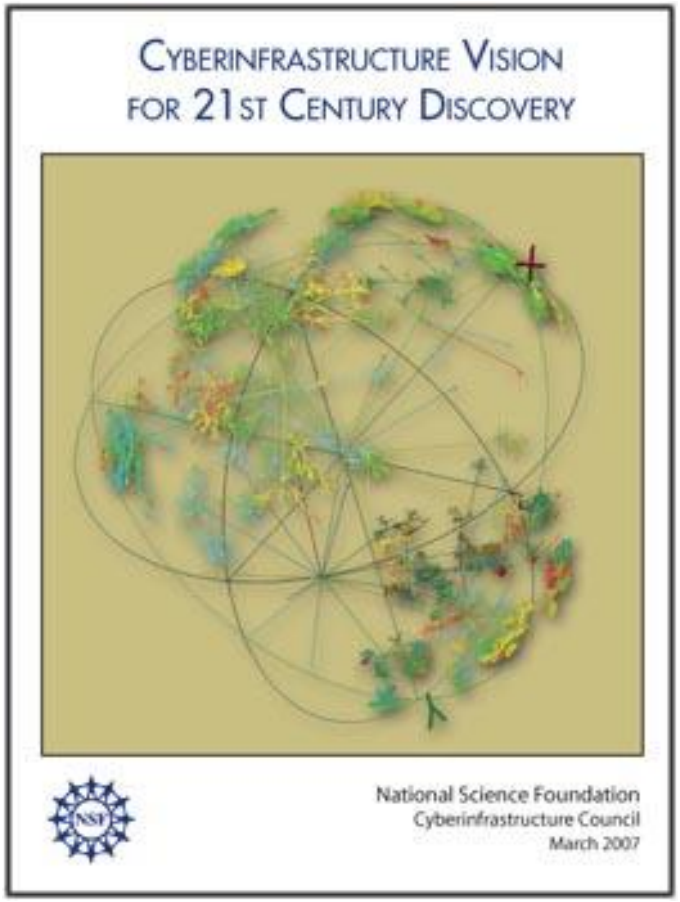
- Research is becoming more data driven.
- Increasingly, data has research value beyond the project it was generated for.
- Data Deluge – the amount of data being produced is increasing exponentially, and the systems, tools and infrastructures to organize, manage, disseminate and preserve data have not kept pace.

Defining Data Curation

- "Data 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." P. Lord, et al.

- "Data curation [is] the value-added activities and features that stewards of digital content engage in to make digital content meaningful or useful." N. McGovern

A Vision

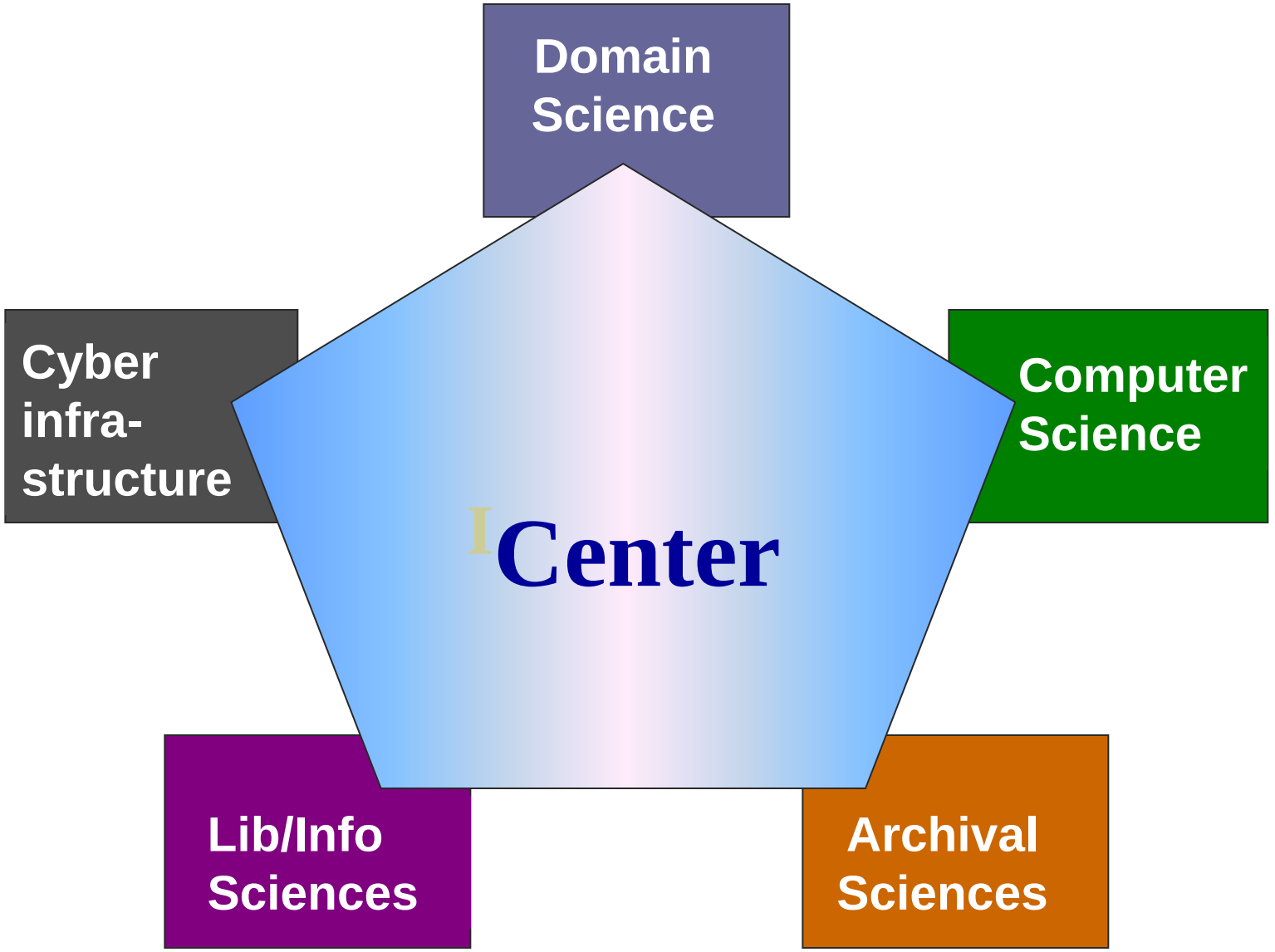


"…science and engineering data are routinely deposited in well-documented form, are regularly and easily consulted and analyzed by specialists and non-specialists alike, are openly accessible while suitably protected, and are reliably preserved…"

Roles for Libraries

"What role do the research and academic libraries envision for themselves and do scientists envision for librarians in a digital data framework…?"

- To Stand the Test of Time…(2006) ARL. p.24



National Science Foundation
WHERE DISCOVERIES BEGIN

Purdue Libraries

Jim Mullins
Dean of the Libraries

- 2004 initiative for Libraries to collaborate with faculty across campus—apply library science knowledge and expertise to research problems: **manage, organize, describe, disseminate, preserve information.**
- Particular emphasis on addressing data curation issues.

Research traditionally moves in this direction

"published" data/ datasets

unpublished research traditional/non

"published" research non-traditional

published research traditional

secondary tertiary resources

analyzed data/ datasets

processed data/ datasets

"raw" data/ datasets

Library service traditionally moves in this direction

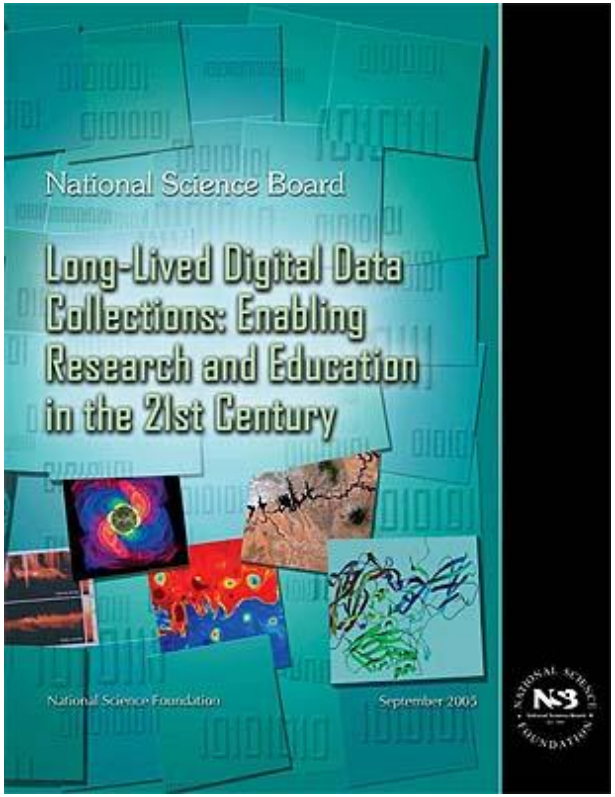
Given the changing nature of research and scholarly communication (i.e., e-research), how do librarians change to adapt to new ideas regarding what they collect to support research & education and how they provide access to it?

Modified from: Brandt, D.S. "Scholarly Communication" (in To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering.: Final Report of Workshop New Collaborative Relationships: Academic Libraries in the Digital Data Universe. ARL, Washington, DC, September 2006.)

The D2C2

- Libraries sponsored research center.
- Established in 2006 to focus on issues associated with curating data sets for present and future research use.
- Working in partnership with domain scientists and IT personnel to address the real world data needs of a research community.

The Data Research Scientist



- Help the Libraries move research strategically forward.
- Help ramp up interaction with research faculty on campus.
- Leverage interdisciplinary research collaborations.
- Address the social, cultural and organizational aspects of data curation.

My Background

- Earned my MLS in 1998.
- Formerly a Subject Librarian for the Social Sciences at a Liberal Arts Institution.
- Skills and experience are mostly those of a “traditional” librarian. Minimal technology background and expertise.

Interacting with Researchers

- Building Connections & Relationships:
  - Investigate
  - Interrogate
  - Invest
- Common Barriers to Data Curation:
  - Motivation
  - Mechanisms
  - Money

Project Types

- Building Capacity
- Addressing Researcher Needs
- Partnerships in Research Projects

Building Capacity

INSTITUTE of Museum and Library SERVICES

PURDUE UNIVERSITY LIBRARIES

Access. Knowledge. Success.

ILLINOIS

- Data Curation Profiles:

  A tool for librarians and others to gather information on researcher needs for their data and to inform curation services.

DCP Sections

- Information about the Data and its Context
  - Overview of the Research
    - Focus
    - Intended Audience
    - Funding
  - Data Kinds and Stages
    - Data Narrative (data lifecycle)
    - Target Data for Sharing
    - Use/re-use Value
    - Contextual Narrative

Data Table – Traffic Flow

Data Stage	Output	Typical File Size	Format	Other / Notes
Primary Data				
Raw	Sensor data	100k in 1 file per day	proprietary to the sensor	FTP downloads are mostly automated.
Processing Stage 1	Sensor data – open/accessible format	Roughly 6kb	.csv / .xls	Data are formatted into .csv before bring reformatted into a mySQL database.
Processed	Data vectors	800 records per intersection per day.	SQL / .xls	Data are extracted from the mySQL database for analysis purposes.
Analyzed	charts/ Graphs		.xls / .emf	charts and graphs used for interpretation.
Published	charts/ graphs		.ppt	Data are presented via power point.
Augmentative Data				
Image	Stills taken from video		.gif /.jpg / .ppt	Images generated from video.

DCP Sections

- Information about Needs

- Intellectual Property
- Organization and Description of Data
- Ingest
- Access
- Discovery
- Tools
- Interoperability
- Measuring Impact
- Data Management
- Preservation

http://datacurationprofiles.org

Addressing Researcher's Needs

Summer Camp 2010

CAMP CALCIUM 11

6 weeks of Camp between June-August 2010

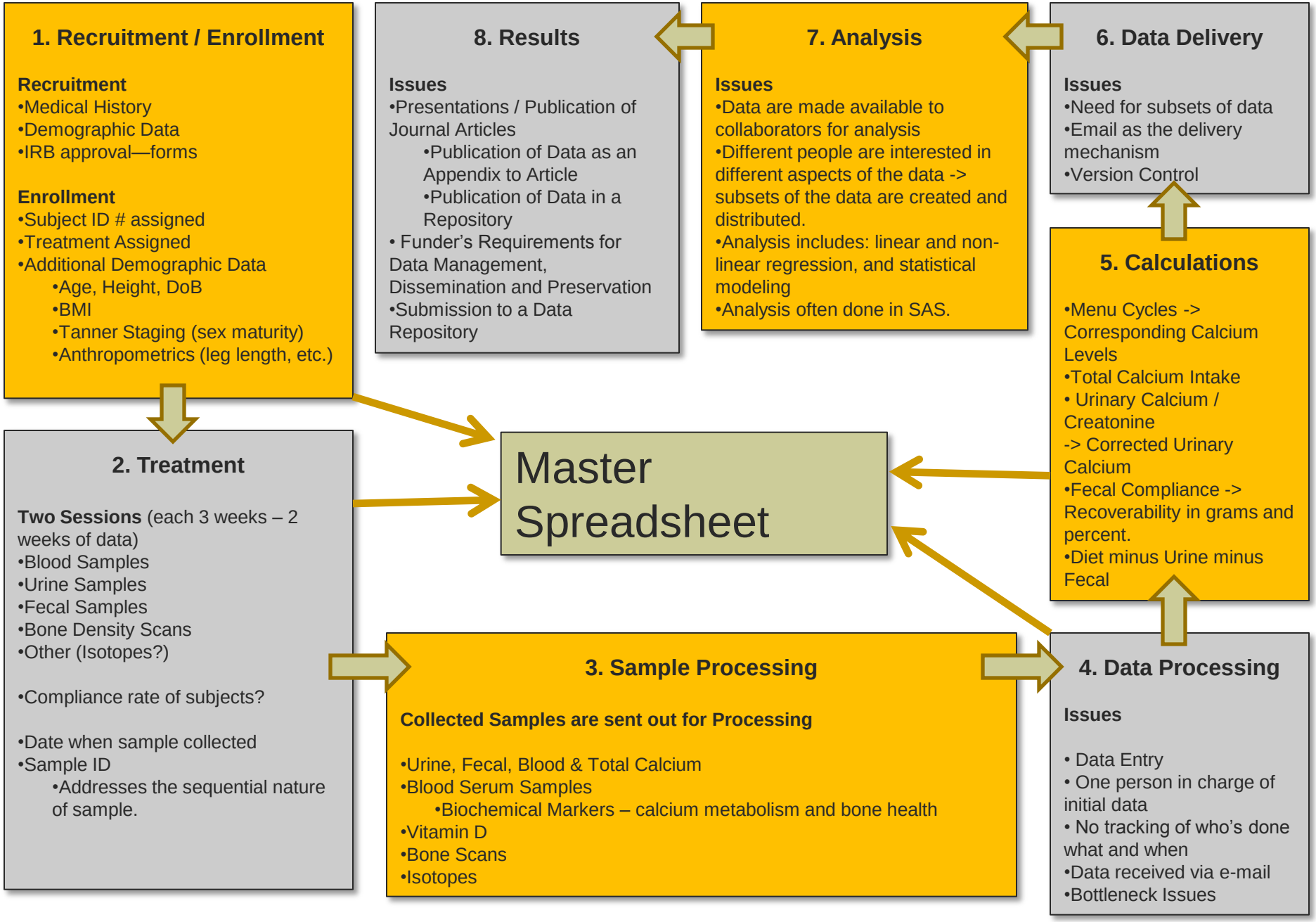
Camp Calcium is:

- For GIRLS 12-14 and BOYS 13-15 of Mexican-American Descent
- A research project. You will learn how much calcium is needed for building strong bones
- 6 weeks! You will live on Purdue campus for two 3-week periods.
- FUN! You will participate in many activities including sports, science classes, nutrition classes, arts and crafts, field trips
- FREE ROOM and BOARD. Registration fee of $100. You will be paid for participation in our research project.

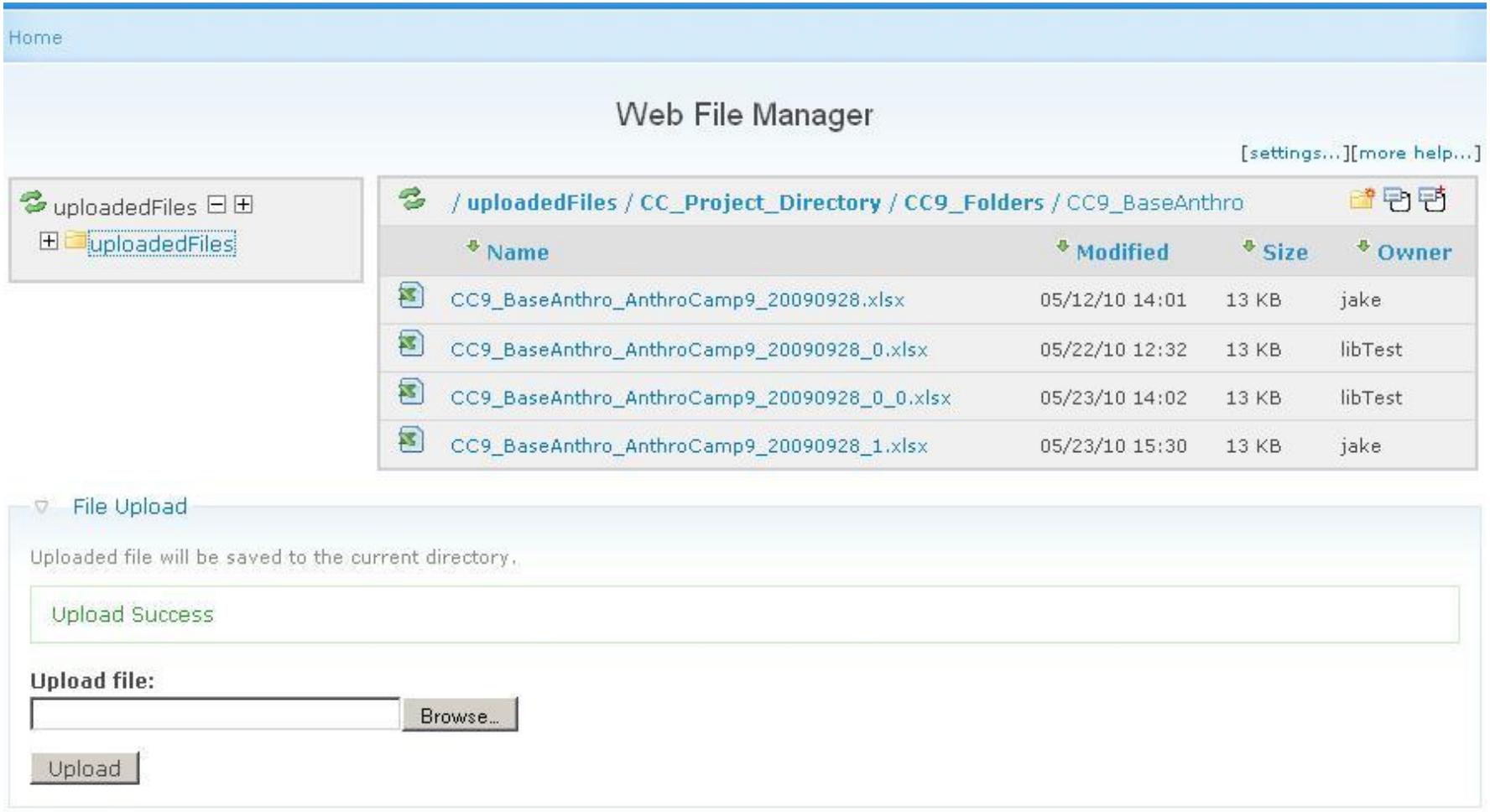
For more information and an application packet contact:

Berdine Martin
700 W. State St.
Purdue University
W. Lafayette, IN 47907-2059
1-800-830-0175 or 765-494-6559
bmartin1@purdue.edu
www.cfs.purdue.edu/fn/campcalcium

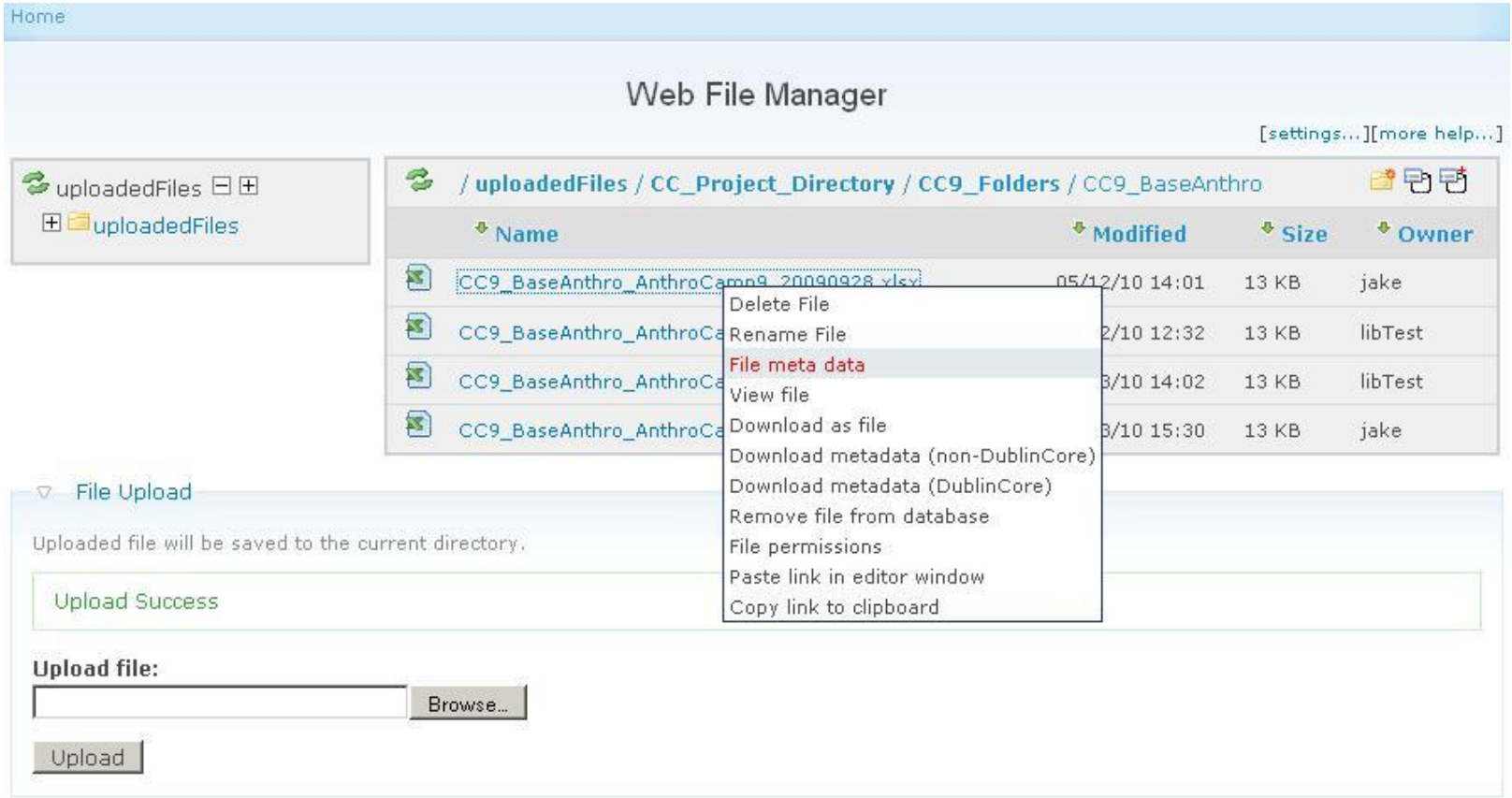
Camp Calcium – Current Data Workflow



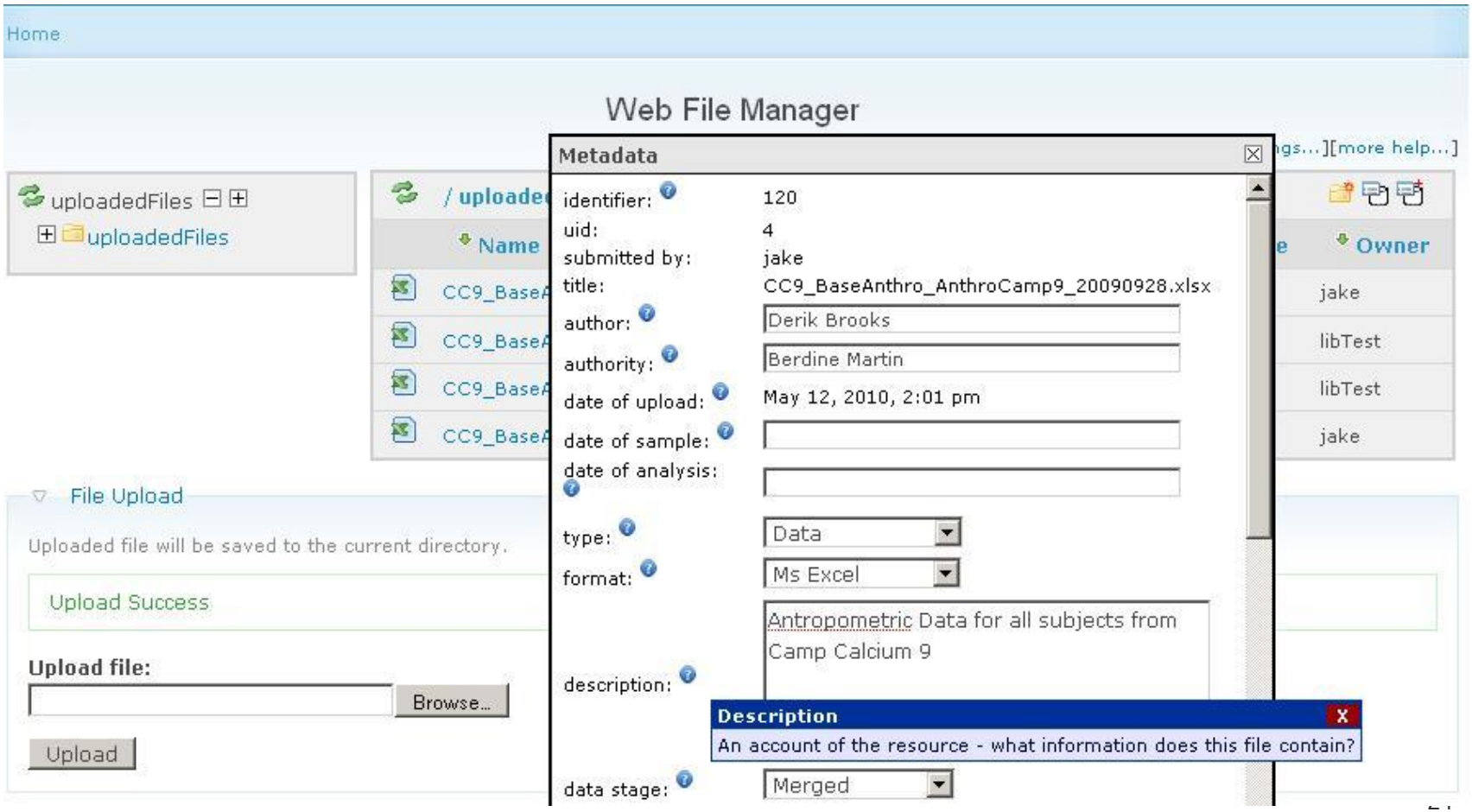
Data Management System



Metadata


Home
Web File Manager
[settings...][more help...]
uploadedFiles
uploadedFiles
/ uploadedFiles / CC_Project_Directory / CC9_Folders / CC9_BaseAnthro
Name
Modified
Size
Owner
CC9_BaseAnthro_AnthroCamp9_20090928.xlsx
05/12/10 14:01
13 KB
jake
CC9_BaseAnthro_AnthroCa
2/10 12:32
13 KB
libTest
CC9_BaseAnthro_AnthroCa
3/10 14:02
13 KB
libTest
CC9_BaseAnthro_AnthroCa
3/10 15:30
13 KB
jake
Delete File
Rename File
File meta data
View file
Download as file
Download metadata (non-DublinCore)
Download metadata (DublinCore)
Remove file from database
File permissions
Paste link in editor window
Copy link to clipboard
File Upload
Uploaded file will be saved to the current directory.
Upload Success
Upload file:
Browse...
Upload

Metadata Fields

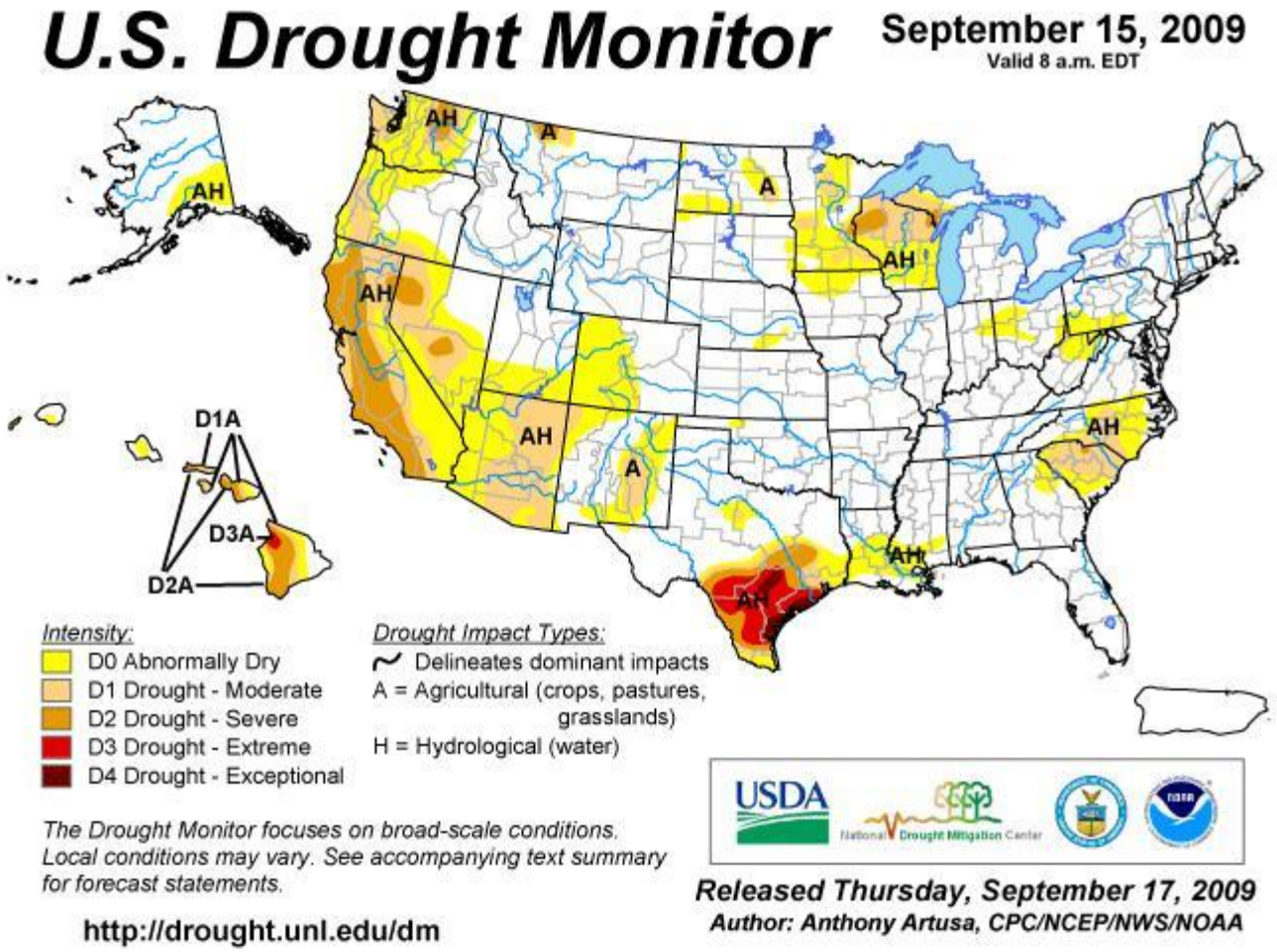


Research Partnerships

- The DRought Information Network (DRInet): Develop a regional scale information web portal and tools for collecting, synthesizing, analyzing and disseminating datasets relating to drought.

National Science Foundation
WHERE DISCOVERIES BEGIN

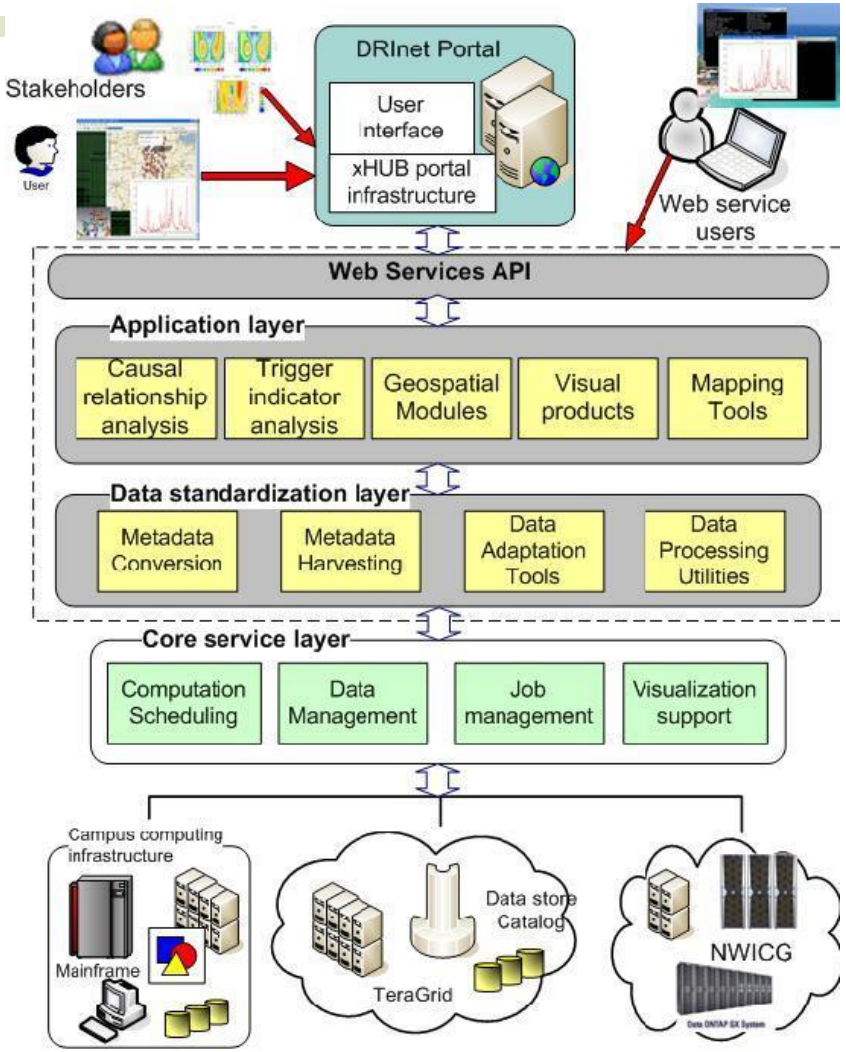
U.S. Drought Monitor



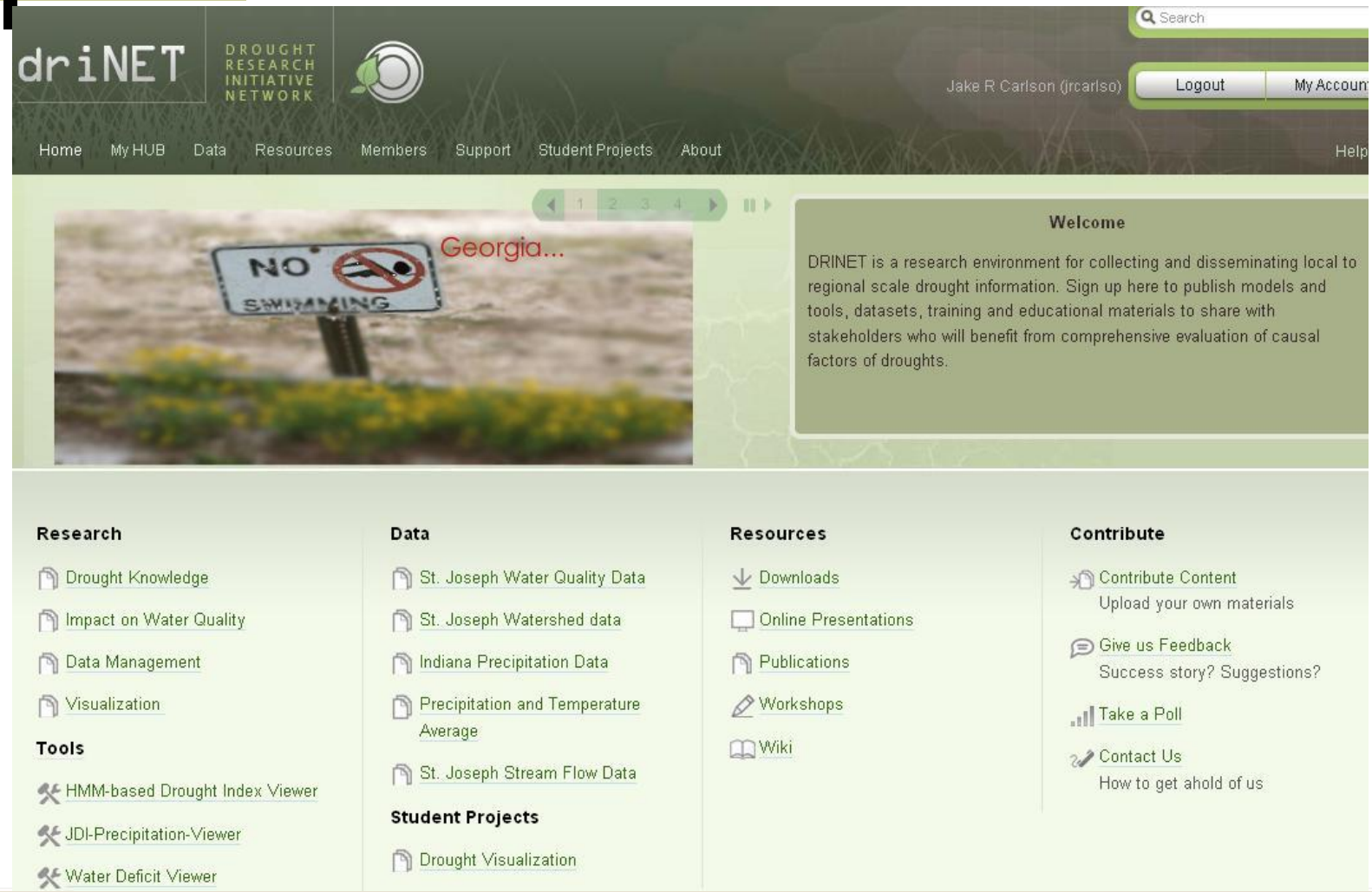
driNET

Role of the Libraries:

- Assist in the identification of metadata standards that will meet user needs.
- Develop discovery and navigation metadata for the portal.
- Assist in the development and testing of tools to automate the metadata creation process.
- Aid in the development and deployment of an interoperability standard (OAI-PMH) to facilitate the dissemination of data.



driNET

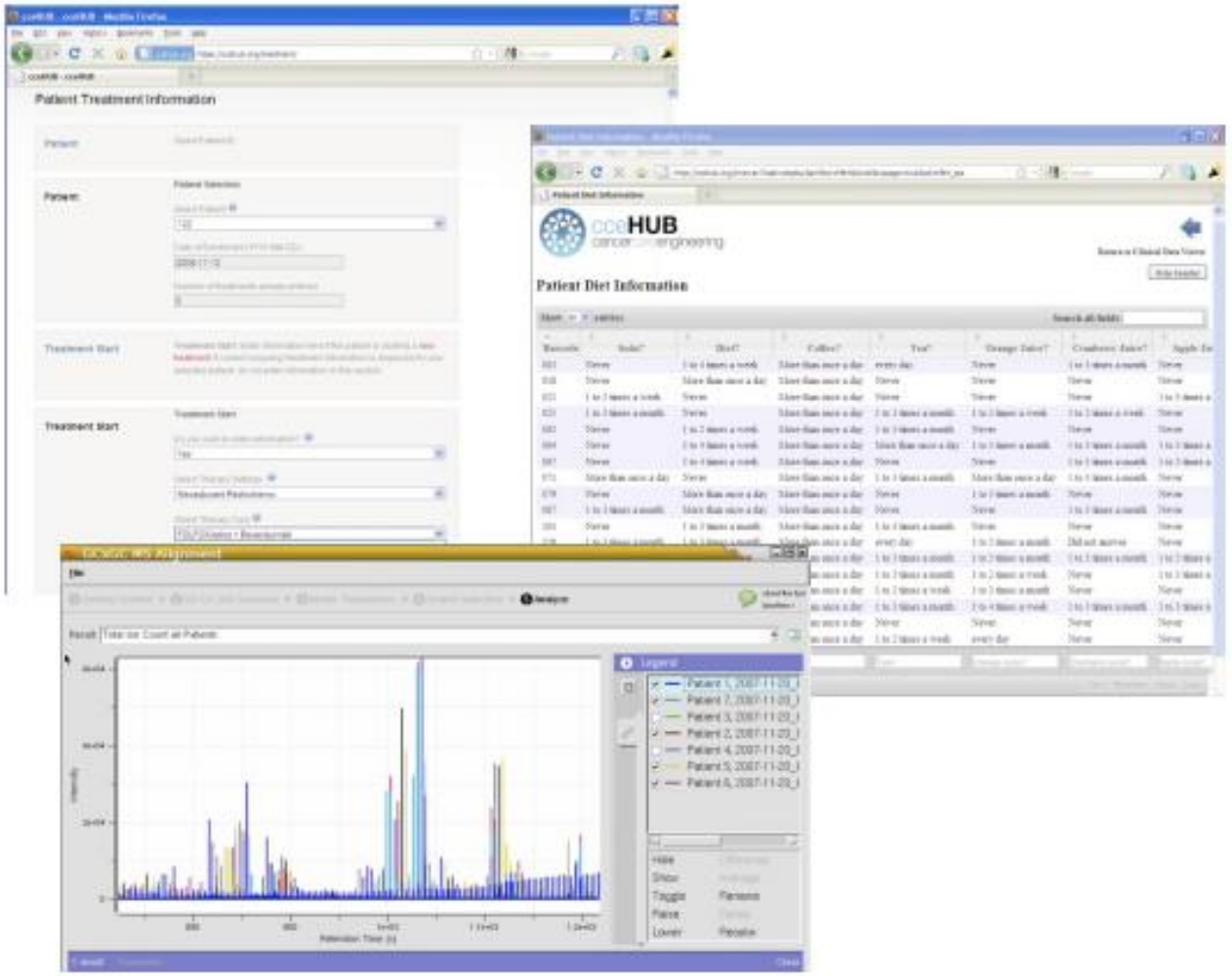

driNET
DROUGHT RESEARCH INITIATIVE NETWORK
Search
Jake R Carlson (jrcarlso)
Logout
My Account
Home
My HUB
Data
Resources
Members
Support
Student Projects
About
Help
Georgia...
NO SWIMMING
Welcome
DRINET is a research environment for collecting and disseminating local to regional scale drought information. Sign up here to publish models and tools, datasets, training and educational materials to share with stakeholders who will benefit from comprehensive evaluation of causal factors of droughts.
Research
Drought Knowledge
Impact on Water Quality
Data Management
Visualization
Tools
HMM-based Drought Index Viewer
JDI-Precipitation-Viewer
Water Deficit Viewer
Data
St. Joseph Water Quality Data
St. Joseph Watershed data
Indiana Precipitation Data
Precipitation and Temperature Average
St. Joseph Stream Flow Data
Student Projects
Drought Visualization
Resources
Downloads
Online Presentations
Publications
Workshops
Wiki
Contribute
Contribute Content
Upload your own materials
Give us Feedback
Success story? Suggestions?
Take a Poll
Contact Us
How to get ahold of us

What's Coming: Increased Interest / Oversight

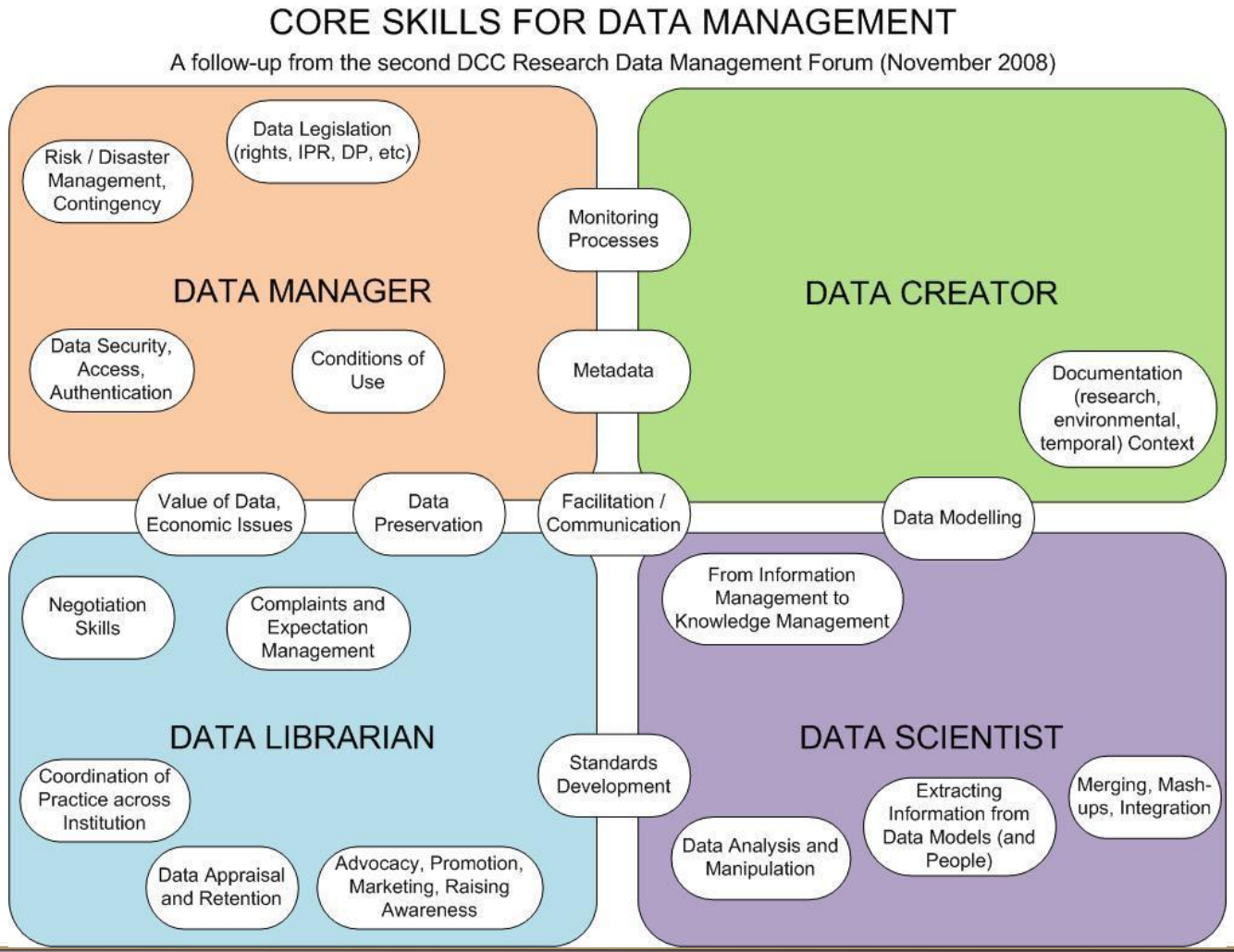
- OVPR's Research Retention Policies and Practices
- H. R. 5116 – "America COMPETES Reauthorization Act of 2010" (April 22, 2010)

(b) Establishment.—The Director of the Office of Science and Technology Policy shall establish a working group under the National Science and Technology Council with the responsibility to coordinate Federal science agency policies related to the dissemination and long-term stewardship of the results of unclassified research, including digital data and peer-reviewed scholarly publications, supported wholly or in part by funding from the Federal science agencies.

What's Coming: Data Repositories



What's Coming: Skills / Roles



Research Data Management Forum: RDMF2: Core Skills Diagram

http://data-forum.blogspot.com/2008/12/rdmf2-core-skills-diagram.html

Thank you!
Questions?

Jake Carlson
Data Research Scientist
Purdue University Libraries & D2C2

Transportation Librarians Roundtable