December 2008

Investigating Data Curation Profiles Across Multiple Research Disciplines

D. Scott Brandt
Purdue University, techman@purdue.edu

Jake R. Carlson
Purdue University, jakecar@umich.edu

Melissa Cragin
University of Illinois at Urbana-Champaign, cragin@illinois.edu.edu

Bryan Heidorn
University of Illinois at Urbana-Champaign

Carole Palmer
University of Illinois at Urbana-Champaign, clpalmer@illinois.edu

See next page for additional authors

Follow this and additional works at: http://docs.lib.purdue.edu/lib_research

Authors

D. Scott Brandt, Jake R. Carlson, Melissa Cragin, Bryan Heidorn, Carole Palmer, Sarah Shreeves, and Michael Witt

Investigating Data Curation Profiles Across Multiple Research Disciplines

D. Scott Brandt[1], Jacob Carlson[1], Melissa Cragin[2], Bryan Heidorn[2], Carole Palmer[2], Sarah Shreeves[2], Michael Witt[1]

[1] Purdue University & [2] University of Illinois, Urbana-Champaign

PROJECT DESCRIPTION

In a two-year project funded by the Institute of Museum and Library Services, investigators from Purdue University and the University of Illinois, Urbana-Champaign are addressing the question, *"Which researchers are willing to share data, when, with whom, and under what conditions?"* With the assistance of graduate students, Marina Kogan and Deborah Leiter, extensive interviews are being conducted with researchers in different disciplines about their scientific workflow and range of research outputs (e.g., raw data through published forms) to identify needs for discovery, access, usage, and re-use of data. The interviews utilize a protocol to elicit responses for cross-discipline comparison and contrast as well as coding to identify categories for deeper analysis. Disciplines covered in the study include:

◊ **Agronomy & Soil Science**
◊ **Anthropology**
◊ **Biochemistry**
◊ **Biology**
◊ **Civil Engineering**
◊ **Earth & Atmospheric Sciences**
◊ **Electrical & Computer Engineering**
◊ **Food Science**
◊ **Geology**
◊ **Horticulture & Plant Science**
◊ **Kinesiology**
◊ **Speech and Hearing**

There are overlaps in three disciplines between Purdue and UIUC, and there are three areas that comprise three or more interviews to generate discipline-specific case studies.

Project goals are to enrich understanding of access to (and sharing of) data and related curation by studying researchers' data practices, translating and comparing needs for archiving and sharing data in curation profiles, and converting the results into formalized policies that can be used by repositories to enhance curation and access to data collections.
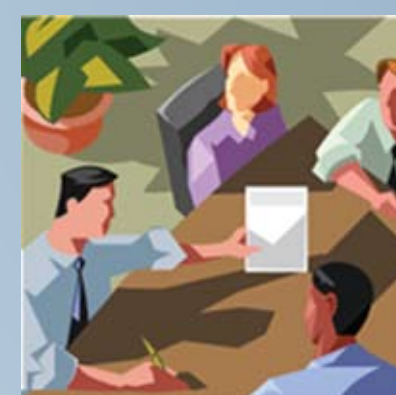
Investigating differences between scientific domains and institutional cultures to inform a general understanding of data curation problems and needs.
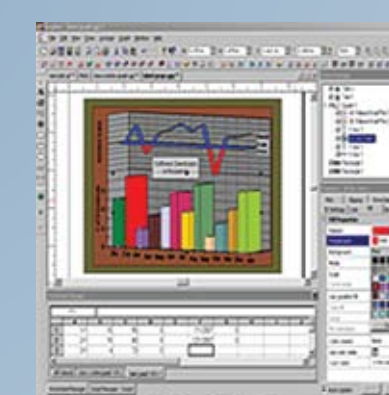
Interviews

Needs & Requirements Analysis

Focus Groups

Data Curation Profiles

Case Studies

Real-World Application

INSTITUTE *of* **Museum** and **Library** SERVICES

METHODOLOGY

1. Interviews with Scientists - Investigators are interviewing scientists and data managers involved with the development of research datasets.

2. Focus Groups With Librarians - Three focus groups will be conducted with the liaison librarians to learn about their work with academic researchers related to data issues and changing roles for librarians vis-à-vis data curation. Many of these librarians facilitated and attended the interviews.

3. Case Studies - Case studies are being developed for two of the disciplines where multiple scientists from the same discipline have been interviewed.

4. Needs & Requirements Analysis - Information from the focus groups and the collected data is being analyzed and distilled into assertions that are made by researchers in roles as data producers, suppliers and users, relative to data preservation and access. Researchers will identify and clarify elements that relate to data producer needs for curation. Common language statements ("I want share data as soon as it is produced") will be translated and correlated to more formal curation activities.

5. Data Curation Profiles - For each case dataset, a curation profile will be created that iterates policy level description into parameters in a manner that allows them to be compared to each other (e.g., as in a matrix) and to be sculpted into a formalized policy language that can be expressed for machine implementation.

6. Real-World Application - An evaluation of current repository systems will be conducted to determine if the capabilities currently exist to meet the needs expressed by the scientists' in the "real world". The ultimate outcome is to determine how this research can be applied to aid researchers in sharing data as appropriate and needed, and to aid librarians in doing so.

For more information visit: **http://datacurationprofiles.org**.