

Centuries of Knowledge  
Data Curation Education Program

IMLS RE-05- 05-0036

University of Illinois at Urbana-Champaign

Interim Report #4  
May 2, 2008 – October 31, 2008

P. Bryan Heidorn, Ph.D., PI  
Graduate School of Library and Information Science  
(217) 244-7792  
[pheidorn@illinois.edu](mailto:pheidorn@illinois.edu)

Melissa Cragin, MLIS  
Project Coordinator  
(217) 244-5574  
[cragin@illinois.edu](mailto:cragin@illinois.edu)

Aaron Collie  
Research Assistant  
(217) 244-5574  
[collie1@illinois.edu](mailto:collie1@illinois.edu)

## **Purpose of the Project**

The primary goal of the Data Curation Education Program (DCEP) is to design a program of graduate study that can serve as a model for training data curators (DCs) within the context of a larger LIS education. Secondly, we intend to integrate this graduate training with ongoing research and practice to produce specialists that understand the research culture and can make substantive contributions to the mission of scientific, humanities, social science, and cultural heritage institutions and libraries.

We are developing a program to train a new generation of LIS professionals qualified as data curators and provide continuing education opportunities for practitioners already in the field. Through workshops, conference presentations, publications, and the influence of our graduates, it will also raise the visibility of the importance of information specialists in managing our knowledge resources for many decades, possibly centuries, to come. These students will become the leaders who build and maintain data systems to work in concert with the many digital libraries, archives, and repositories, as well as the indexing systems, metadata standards, ontologies, taxonomies, and vocabularies associated with digital data and products. In addition, in response to frequent calls from GSLIS alumni and other practicing LIS professionals for data curation training, we have developed the Summer Institute on Data Curation. Details on this inaugural session are included below.

During this last period of the project our work focused on the following areas:

1. Preparing for and conducting the 2008 Summer Institute on Data Curation
2. Needs assessment
3. Recruitment of students
4. Development of internship opportunities
5. Working with our Advisory Committee
6. Coordination and dissemination activities
7. Curriculum development (including development and implementation of new required courses)

## **Activities and services carried out with project funds during this reporting period:**

### 1. 2008 Data Curation Summer Institute

In our October, 2007 Interim Report we described the ongoing requests that practicing librarians have made to DCEP faculty and staff for training or classes in data curation. To address this need, we developed and delivered a 4-day Summer Institute on Data Curation, June 2-5, on the UIUC campus. Thirty people attended the Institute. The Institute program included the following presentations:

- Introductions and Overview of the week - Melissa Cragin and Carole Palmer (GSLIS)
- Introduction to Data Curation and general issues - Melissa Cragin
- Introduction to Digital Data - W. John MacMullen (GSLIS)
- Integrity and Authenticity - Allen Renear (GSLIS)
- Techniques for Evaluating Scientific Data Practices and Needs - Carole Palmer
- Appraisal and Selection - Melissa Cragin
- Preparing Data for Ingest - Ruth Duerr (National Snow and Ice Data Center)
- Resource Requirements for a Data Curation Program - Scott Brandt (Purdue University Libraries)
- Panel: Librarians and Scientists Working Together
- Day-to-Day Digital Preservation - Tim Donohue (UIUC Libraries)
- Digital Preservation and Standards - Jerome McDonough (GSLIS)

Digital Preservation and Standards - Hands-On - Jerome McDonough (GSLIS)  
Technical Aspects of Digital Library and Repository Systems, Architectures and Infrastructures -  
Tim DiLauro (Johns Hopkins Libraries)  
Keynote address - Anna Gold (Kennedy Library, California Polytechnic State University)

Following is a list of institutions represented at the 2008 Summer Institute, which included 13 ARL Member organizations:

Canada Institute for Scientific and Technical Information (CISTI) –3  
Columbia University Libraries – 1  
University of Pennsylvania Library – 2  
Georgia Institute of Technology – 1  
Northwestern University Library – 1  
Ohio State University Libraries – 1  
Purdue University Libraries – 3  
University of California, San Diego Libraries – 1  
University of Illinois at Urbana-Champaign Library – 3  
University of Minnesota Libraries – 3  
University of Virginia Library – 2  
University of Washington Libraries – 2  
University of Wisconsin – 3 (two librarians, one from University IT)  
Claremont University Consortium –1  
Stanford University Library – 1  
Woods Hole Oceanographic Institution –1  
University of Illinois at Urbana-Champaign GSLIS Bioinformatics Student - 1

## 2. Needs Assessment

We have finished the data collection for the first Needs Assessment, on which we collaborated with the Environmental Council (EC) at the University of Illinois at Urbana-Champaign. As the return rate from the web-based survey was low, we modified our method and called those in our pool who had not yet completed the survey. We have started to analyze the data, and will include highlights in our next report. In addition, we will develop the next Needs Assessment study over the coming two months; we intend to redevelop the survey instrument itself based on analysis of several related surveys, and expand our study population across academic institutions.

## 3. Student Recruitment

We have just conducted another mailing of recruiting materials to undergraduate science and computer science programs in Illinois. Current recruiting mail list includes:

- a. All undergraduate programs in biology, geology, geography, environmental sciences, and anthropology in Illinois
- b. McNair Scholars Program sites in selected colleges and universities throughout the United States
- c. Colleges and universities with undergraduate library and information science programs

In addition to the list above, we are now compiling a list of science departments at the Historically Black Colleges and Universities, in order to expand a possible pool of students from underserved populations.

During the 2007-2008 academic year, 5 students enrolled in the program. One graduated in August, 2008, and two more will graduate in December, 2008. Beginning this current academic year,

2008-2009, we have enrolled 12 new students into the DC concentration. In the past we had awarded a year-long fellowship to one of the new campus students, and are awarding another fellowship to one of the new LEEP online students for the spring and summer 2009 semesters.

#### 4. Internship Opportunity Development

Internship opportunities for our students are under development at several institutions, including the Smithsonian Institution, Purdue University Libraries, and the National Library of Medicine. The GSLIS Bioinformatics program is benefiting from our intern placement efforts; as there were not yet any DC students prepared for fieldwork opportunities, in summer, 2008 we placed a bioinformatics student in a summer-long curation internship sponsored by the Johns Hopkins University Libraries and Computer Science Department. As we will have DC students preparing for graduation in the next academic year, we fully anticipate matching them with curation internship opportunities currently being established.

In addition, we have recently requested, and have approval from IMLS, to broaden the source of candidates for internships in data curation to include students in the GSLIS Certificate of Advanced Study and the GSLIS Master of Science in Bioinformatics programs. This request was initiated because we are finding that two other groups of GSLIS students are enrolling in courses in the Data Curation (DC) curriculum, though they are not enrolled in the Master of Science degree program. First and foremost, there are CAS (Certificate of Advanced Study) students who are now taking our DC courses. These students already have MSLIS degrees and are continuing their education with additional coursework. Many of these students have a level of technical expertise that make them a great match for the internship opportunities that we are developing as part of the GSLIS portfolio of field placement opportunities. The other group of students are enrolled in the GSLIS program for the University of Illinois Master of Science in Bioinformatics. These students enter the program with a degree in science, and some with computer science as well. Several of these students are now taking the core courses in the Data Curation concentration, and they, too, might be a great “fit” for some of the more computationally demanding internships. Because these students are participating in the DC curriculum, we would like to be able to support them in educational and career development opportunities in this area. Not only will this benefit the students, but by widening the intern pool, we will be in a better position to secure and maintain field placement opportunities as the Data Curation field grows.

#### 5. Advisory Committee Activities

On November 11, 2008 we will hold our fourth Advisory Committee Meeting, via teleconference. We will report on the curriculum development, and the new Data Curation in the Humanities grant (described at the end of this report). We will also spend some time reviewing the progress of the case studies that the Advisors are helping us to develop.

The Advisory Committee members currently include:

- ◇ Chris Freeland, Application Development Manager and Project Manager, Missouri Botanical Garden
- ◇ Thomas Garnett, Associate Director for Digital Library and Information Systems, Smithsonian Institution
- ◇ Martin Kalfatovic, Head of the New Media Office and Preservation Services Department, Smithsonian Institution Libraries, Washington, D.C.
- ◇ Joanna McCaffrey, Collections Database Architect, The Field Museum
- ◇ Nancy McGovern, Digital Preservation Office, Inter-University Consortium for Political and Social Research (ICPSR), University of Michigan
- ◇ Katherine McNeill-Harman, Data Services and Economics Librarian, Massachusetts Institute of Technology

- ◇ Chuck Miller, Vice President, Information Technology and Chief Information Officer, Missouri Botanical Garden, St. Louis, MO
- ◇ David Soller, Ph.D., Geologist, U.S. Geological Survey, Reston, VA

## 6. Coordination and Dissemination Activities

The website was redesigned, and is located at <http://www.uiuc.edu/goto/dcep>. Materials developed *about* the DCEP and dissemination activities are listed on the project web site; the materials are also deposited into the UIUC institutional repository, the Illinois Digital Environment for Access to Learning and Scholarship (IDEALS).

Ellen Rubenstein, the Graduate Assistant (GA) for the 2007-2008 academic year, worked on most of the activities described above, with particular focus on recruitment, the Needs Assessment, and planning for the Summer Institute for Data Curation. Our new Graduate Assistant is Aaron Collie, who began work on August 16, 2008. To date his focus has been on recruitment, tracking Data Curation (and related) job postings, and the redevelopment of the DCEP web site.

## 7. Curriculum Development

We have recently identified a gap in our curriculum, and are actively working to hire an adjunct to help us develop a course on Data Architectures and Curation Tools. This course will address theoretical, technological, human factors and evaluative components of current and emerging technical infrastructures for data.

The Foundations of Data Curation course is being offered for the second time this semester, taught by Prof. W. John MacMullen. Enrollment for the course includes 14 GSLIS MS students, 3 Community Credit students (2 PhDs, 1 practicing academic science librarian), 2 CAS students, 1 WISE MS student 1 bioinformatics MS, 1 auditor. During this current academic year, we will apply to the Graduate College for a permanent course number.

Digital Preservation will be offered again in the spring of 2009, taught by Prof. Jerome McDonough. Systems Analysis and Management will also be taught in the spring semester.

As noted in previous reports, DCEP course materials, such as the syllabi and case studies will be made available through the DCEP web site, as well as the University of Illinois IDEALS institutional repository.

**The following list of presentations represent those that occurred during this reporting period:**

### Invited talks

Cragin, Melissa H. "Data Curation Services in Research Libraries." National Library of Medicine, April 15, 2008.

### Panels and Workshops

Palmer, Carole L. (2008). E-research crosses the pond: Contrasting transformations in the U.S. and U.K. Presentation accepted for panel organized by Eric Meyer, American Society for Information Science and Technology Annual Meeting (ASIS&T), October 24-29, 2008, Columbus, Ohio.

Palmer, Carole L. Developing curriculum for digital libraries and digital curation education: Reflections on synergies and divergencies. Panel presentation organized by Jeffrey Pomerantz, et al., American Society for Information Science and Technology Annual Meeting (ASIS&T), October 24-29, 2008, Columbus, Ohio.

Palmer, Carole L. (2008). Data curation education: Evolving research librarianship in the sciences. Librarians & e-Science: Focusing towards 20/20. Committee on Institutional Cooperation, Center for Library Initiatives Conference, May 13, 2008, West Lafayette, Indiana.

Smith, Linda C., Cragin, M.H., Heidorn, P. B., Palmer, C.L., MacMullen, W. J. (2008). Developing an Effective Data Curation Education Program. Presentation at the "Education for Digital Stewardship: Librarians, Archivists or Curators?" workshop at the Joint Conference on Digital Libraries (JCDL), June 20, 2008; Pittsburgh, PA.

### Posters

MacMullen, W. John. Searching for Uses and Users in Gene Ontology Research. Short paper & poster for the Annual Meeting of the American Society for Information Science & Technology (ASIS&T), October 24-29, 2008, Columbus, Ohio.

MacMullen, W. John (2008). Understanding biocurators: Attributes and roles of model organism database curators. Poster for the Medical Library Association (MLA) Annual Meeting, May 16-21, 2008, Chicago, Illinois.

### Other Activities

Palmer, Carole L. Education for new roles. Presentation at an Association for Research Libraries Fall Forum, Reinventing Science Librarianship: Models for the Future. October 16-17, 2008. Washington, D.C.

Palmer, Carole L. (2008). Research practice and research libraries: Working toward high-impact information services. OCLC Distinguished Seminar, June 18, 2008, Dublin, Ohio.

### Service Activities

P. Bryan Heidorn, Ph.D. (PI) is now Program Director, Division of Biological Infrastructure, National Science Foundation. He serves on the e-Biosphere Steering Committee, and continues his work on the Board of Directors of the JRS Biodiversity Foundation.

Melissa Cragin is serving on the Program Committee for the Fourth International. Digital Curation Conference, to be held in Edinburgh, Scotland, Dec. 1-3, 2008.

### Related Work

Our research project with the Purdue University Libraries, "Investigating Data Curation Profiles across Research Disciplines," is ongoing. We continue to collect data across both the Purdue and UIUC campuses, with particular focus on disciplinary differences in research workflows and data practices, and their implications for data curation services. This study includes a focus on the ways librarians can interact with and support scientists in making their research output available, and the implications for LIS education and in-service training.

We are working directly with the PI and staff of the recently funded IMLS project, "Extending Data Curation to the Humanities: Curriculum Development & Recruiting" led by Assoc. Dean for Research, Allen Renear. The new project will focus on fundamental principles and best practices for managing the entire lifecycle of humanities data: creation, management, exploitation, enhancement, and preservation, developing a model curriculum to share with the field.