Preparing e-Science Information Specialists: New Programs and Professionals

Carole L. Palmer
Center for Informatics Research in Science and Scholarship (CIRSS)

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

ARL / CNI Fall Forum
Reinventing Science Librarianship: Models for the Future
17 October 2008
Three approaches for meeting demands arising in academic libraries and research centers:

- **Biological Information Specialist** in campus bioinformatics program.

- **Data curation concentration** within the MSLIS at GSLIS.

- **Summer institutes** for practicing librarians and information service providers.

Motivation
- Historical and disciplinary context
- Involvement of scientific communities
- Reciprocity with current research
Demands on the scientific workforce

Experiences studying and collaborating with scientists, documenting information requirements and expertise to advance research -

Motivated 2005 NSF proposal argued for information specialists to allow

- scientists to concentrate on scientific problems
- computer scientists to let go of projects as they move into implementation.

Need for professionals to bridge growing universe of

- information resources,
- informatics tools,
- scholarly communication options.
Not so new, new world of scientific information

- Contemporary themes in landmark meetings:
  1948 - Royal Society Scientific Information Conference
  1952 - Chicago School symposium on special information
  1958 - International Conference on Scientific Information

  interrelations of information system, complexity of formats, prepublication document components, speed of circulation, interdisciplinarity…

- 1980s predictions of revolution in scholarly information processing—functionality possible, assumed technologically feasible.

  discipline-specific browsing & analysis, data-driven user-modifiable diagrams, computationally available equations & chemical formulae, structured annotations, thorough-going interoperability…(Coombs, Renear, & DeRose, 1987; Renear & Bilder, 1993)

Professional perspective and priorities

Recognition that applications of information in the practice of science more important than volume and format of information.

Application and further development of grounding theories:

- adding value to improve use potential  (Taylor, 1986).
- coordinating and integrating information in alignment with complex social structures and practices  (Shera, 1972).

The true essence of librarianship...is the maximization of the effective use of graphic records for any purpose….  (Shera, 1971, p. 57).

Next-generation information professionals will build and maintain data and information systems specialized for specific sciences.

AND, MORE IMPORTANTLY...
Coordinate across sciences

- Metascience responsibilities of LIS and research libraries
  - Coordinate landscape of information—federated collections, indexing, metadata standards, integrative ontologies.
  - Assure new systems work in concert with existing and emerging digital libraries, archives, and repositories.

- Draw on foundations in user communities, interoperability, digital preservation, data modeling, ontology development, digital aggregation, information architecture, sustainability.

- Infrastructure and services necessary to facilitate interdisciplinary & multi-scale science require expertise

IN THE LAB & IN THE LIBRARY
Greatest role for research library in small science

Data from *Big Science* is … easier to handle, understand and archive. *Small Science* is horribly heterogeneous and far more vast. In time *Small Science* will generate 2-3 times more data than Big Science.

(‘Lost in a Sea of Science Data’ S.Carlson, The Chronicle of Higher Education, 23/06/2006.)
Essential to smaller, multi-disciplinary science

Data needs assessment of UIUC “Faculty of the Environment”; daunting to define, reach, respond to the user community.
BIS part of campus-wide bioinformatics program (NSF/IIS/CISE – 0534567, 2006, Palmer, PI)

Not focused on computational molecular biology, but biological informatics broadly construed:

“Tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.” (NIH, BISTI)

Began with emphasis on data curation and integrative science. Extension of 30-yr-old informationist (or ISIC) movement.

Clinical medical librarians improve information use and communication as members of scientific research groups.

Requires expertise in LIS and research domain.
Campus core requirement in biology, CS, & bioinformatics.

GSLIS distribution requirement of one course in 3 of 4 areas:
- Information Organization and Knowledge Representation
- Information Resources, Uses and Users
- Information Systems and Access
- Disciplinary Focus

**New course offerings:**
- Introduction to Biological informatics Tools & Resources
- Biodiversity & Ecoinformatics
- Information Transfer & Collaboration in Science
- Literature-based discovery
- Data mining

Thesis strongly recommended
Data curation concentration in MSLIS

Data Curation Education Program (IMLS/LB, 2006, Heidorn, PI)

- Curriculum (distance option), recruitment, field work, needs assessment
  - Digital data collection & management, representation, preservation, archiving, standards, policy.
  - Enabling data discovery and retrieval, maintain quality, add value, and provide for re-use over time.

Required Core Courses

- Foundations of Data Curation
- Digital Preservation
- Systems Analysis & Management
- Field experience seminar

Selected Electives – require 2, recommend 4

- Information Modeling
- Metadata in Theory & Practice
- Ontologies in Natural Science
- Foundations of Information Processing
- Digital Libraries: Research & Practice
- Biodiversity Informatics
- Representing & Organizing Info Resources
Core curation content

**Foundations of Data Curation**
- Digital Data
- Scholarly Communication
- Lifecycles
- Collections
- Infrastructures & Repositories
- Selection and Appraisal
- Metadata
- Standards & Protocols
- Archiving & Preservation
- Intellectual Property & Legal Issues
- Workflows; Data Re-use & Value
- Policy & Cooperative Alignments
- Scientific Information Work

**Assignments:**
- 20 cases developed this semester
- Critiques of data management plans

**Digital Preservation**
- Archival Theory & Diplomatics
- OAIS Reference Model
- Data Formats
- Digital Archival Objects
- Preservation Strategies:
  - Emulation vs. Migration
  - Authenticity, Integrity & Trust
- Evaluation & Value
- Digital Preservation & The Law

**Assignments:**
- Planning Grant Application
- Trusted Repository Assessment
BIS student profiles

Began Fall 2006

- Most have biology degrees, many with masters, some PhDs. Some undergraduate biology degrees coupled with CS minor.

- First graduate in 2007 placed at Kansas State Medical Center – support for medical research statistical & computational data analysis.

- 8 declared students in progress. More than any other Illinois bioinformatics program.

- Many LIS students taking advantage of BIS courses. (particularly from health sciences librarianship core)

- Losing some to LIS due to CS requirements and financial aid.
Data curation student profiles

Began Fall 2007

- 3 graduates as of December.
- 21 currently enrolled, mostly through LEEP distance option.
- Mix of full-time and part-time. Large number of part-time students in sciences entering through community credit route.
- Many currently working: USGS, National Archives, health sciences & ag/environ librarians, AUL for info tech, NCSA research scientist, university IT
- More internship openings than students: NLM, USDA, MBOT, Purdue
- Demand from practicing academic libraries
30 participants

26 from ARL institutions - academic librarians, administrators, IT staff, including 3 from CISTI

10 presenters (GSLIS; National Snow and Ice Data Center; Purdue, UIUC, Cal Poly, and Johns Hopkins Univ. Libraries)

6-person panel (3 librarians and 3 scientists)

Topic areas covered included:

- Digital data
- Data integrity & authenticity
- Appraisal and selection
- Preparation for ingest
- Digital preservation standards
- Day-to-day preservation work
- Repository architectures
Partnerships with premier science data centers

Advisors, instructors, internship sites, use cases & best practices:

- BIRN (Biomedical Informatics Research Network) - Maryann Martone
- Smithsonian Institution Libraries, Biodiversity Heritage Library – Thomas Garnett & Martin Kalfatovic
- U.S. Geological Survey - David Soller
- Marine Biological Laboratory - Indra Neil Sarkar
- Missouri Botanical Garden - Chris Freeland
- Field Museum of Natural History - Joanna McCaffrey
- US Army ERDC-CERL - General William D. Goran
- Snow and Ice Data Center - Ruth Duerr
- Johns Hopkins Libraries – Sayeed Choudhury – 1st Internship placement
Educational programs integrated with CIRSS research

Scientific research practices

Datasets as research collections

Flickr user: stacina creative commons

Flickr user: r h creative commons
Profiling curation requirements across sciences

Curation Profiles Project (IMLS NLG 2007-2009)
Led by Purdue University Libraries (D. Scott Brandt, PI)

- In collaboration with librarians, working closely with scientists to study
  - research data management / metadata workflow
  - policies for archiving and access
  - system requirements for managing data in a repository
  - librarians roles and skill sets to support archiving and sharing
Other projects in e-science area

Bryan Heidorn & collaborators in biodiversity informatics:

- HerbIS: Erudite Recorded Botanical Information Synthesizer.
- Networked Environmental Sonic-Toolkits for Exploratory Research

John MacMullen in genomics:

- Information Integration Using Annotation Evidence

Melissa Cragin in neuroscience:

- Roles of shared digital data collections: A case study in Neuroscience
Research & education programs aims & activities

Provide base for research librarianship

- *Purposeful curation* and services for a future with “working” data

Address fundamental research problems

- Investigate how to analyze & represent the *analytical potential* of datasets

Partner with community to promote development of profession.

- Hosting Digital Curation Centre’s 6th International Conference in 2010
Going forward, important not to underestimate challenge

We are, as a profession, well positioned—with compatible institutional and human infrastructure, expertise, commitment…

And, have made great headway on mastering the bibliographic universe, even as it has evolved.

But that professional knowledge began generations ago.

Studies of scientific information use emerged in 1960s, began in earnest in 1980s. But, first text books did not appeared until recently as base of knowledge organizational and functional requirements matured.

Just beginning to investigate, document, and respond to the much more complex collection and service environment for e-research and data-centric science.
Questions & comments welcome

clpalmer@uiuc.edu

http://cirss.lis.uiuc.edu/