



Data Curation Education: Evolving Research Librarianship in the Sciences

Carole L. Palmer

Director, Center for Informatics Research
in Science and Scholarship (CIRSS)



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

Librarians & e-Science: Focusing towards 20/20
Purdue University
13 May 2008



Philosophy & motivation

Teaching aim - to graduate leaders in LIS:

- Began as topic in Current Topics in Collection Development (1999)
 - You *will* be responsible for “data”

Research on how to advance the work of scientists:

- Information and Discovery in Neuroscience (NSF/CISE 2002-2005)
 - Data management and ontology development most arduous;
 - Lack of adequate work force
- Biodiversity informatics research (IMLS, NSF, Moore 2002-present)
 - Description standards, georeferencing specimen sources, collaborative field data collection and validation





Reestablishing professional roles in science

- **Biological Information Specialists (BIS)** - Part of campus bioinformatics (NSF/IIS/CISE – 0534567, 2006, Palmer, PI)

Informationists for contemporary biology; broad based science generalist with strong data curation emphasis from beginning.

- **Data curation concentration in MSLIS**

(IMLS – Laura Bush 21st Century Librarian Program
RE-05-06-0036-06, 2006, Heidorn, PI)

Focus on digital data collection and management, representation, preservation, archiving, standards, and policy.

Develop curriculum, internships, promote & share DC expertise.





Collaboration with premier science data centers

- Maryann Martone, BIRN (Biomedical Informatics Research Network)
- Chris Freeland & Chuck Miller, Missouri Botanical Garden
- Thomas Garnett & Martin Kalfatovic, Smithsonian Institution Libraries, Biodiversity Heritage Library
- Joanna McCaffrey, Field Museum of Natural History
- David Soller, U.S. Geological Survey
- Indra Neil Sarkar, Marine Biological Laboratory
- Gen. William D. Goran, US Army ERDC-CERL





Extending library functions to new content

The active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education.

Activities

- enable data discovery and retrieval
- maintain data quality
- add value
- provide for re-use over time
- archiving
- preservation

Tasks

- appraisal and selection
- representation
- authentication
- data integrity
- maintaining links
- format conversions





What's new for libraries and librarians?

- Closer engagement with scientists during research production,
 - More sophisticated understanding of the differences in research cultures across domains
 - Potential for more direct contributions to scientific
- Facilitation of data deposition
 - Local, disciplinary, larger federations
- New collaborations and constituencies
 - Campus IT, research officers
- Development of data curation principles and systematic practices





Complexities of data collections

Primary research data, generated in digital form;

Secondary data, mobilized to produce new primary research;

And their various transformations

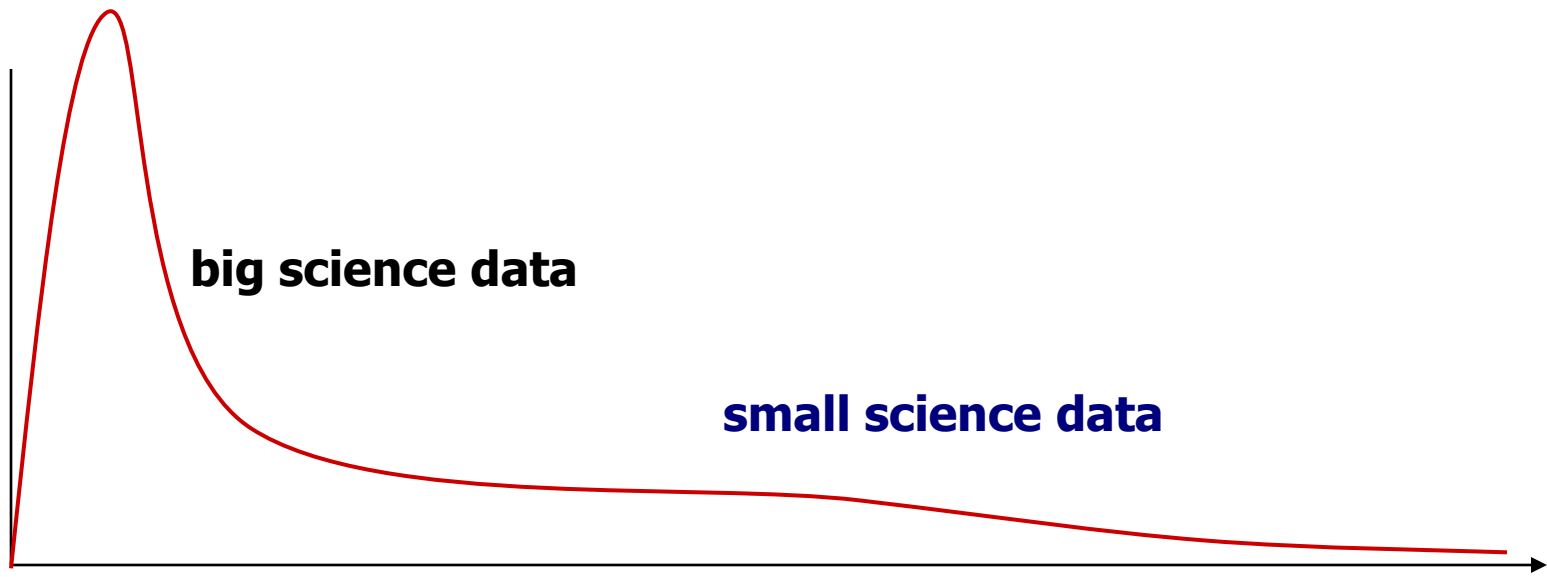
Data Characteristics – Crystallography	
Type	<ol style="list-style-type: none">1. “Raw data” – binary image frames2. “Phased file” – electron density3. “Integrated data” – amplitudes of molecules4. “Corrected data” – according to theory
Format	<ol style="list-style-type: none">1. Binary diffraction images based on the software2. Different electron density image3. Multiple formats4. CIF file
Size	<ol style="list-style-type: none">1. About 2,400 frames $\frac{1}{4}$ -1Mb each – about/over 1Gb2. > 100Mb3. 5-6 Mb4. < 1 Mb
Workflow	well-defined stages, for measurement or analytical purposes, in sequence; output of one stage constitutes the input to the next; for publication CIF considered final result of experiment



Stewardship of small, “special” collections

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.)

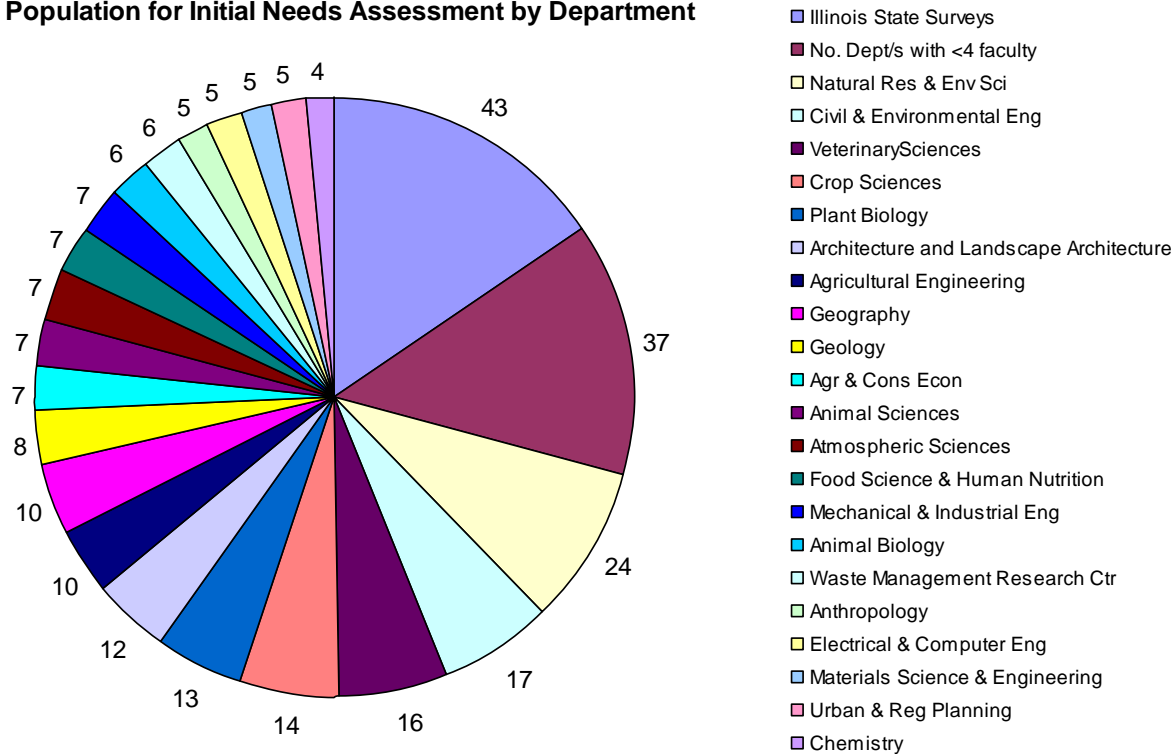




Small science data is scatter and isolated

UIUC faculty of the environment

Faculty Population for Initial Needs Assessment by Department





Curation roles within and across science

In the tradition of research librarianship, professionals must understand the landscape of research resources and how resources work together:

- Collect and manage data in ways that add value
and
promote sharing and integration across laboratories, institutions, and fields of research
- Build and maintain data systems that work in concert with digital libraries, archives, and repositories,
and
the indexing systems, metadata standards, ontologies, etc. associated with digital data and products.





Data curation curriculum

- Required core courses

Foundations of Data Curation

Digital Preservation

Systems Analysis and Management

Field experience seminar

- Selected Electives – 2 required, 4 recommended

Information Modeling

Metadata in Theory and Practice

Ontologies in Natural Science

Foundations of Information Processing

Digital Libraries: Research & Practice

Biodiversity Informatics

Representing and Organizing Information Resources





Core course content

Foundations of Data Curation

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

Digital Preservation

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

Assignments on:

- Planning Grant Application
- Trusted Repository Assessment





Demand and related challenges

- More internship openings than students
 - Data curation and BIS students proving to be attractive to research centers and library operations
- Expressed need for continuing education
 - 1st summer institute for academic librarians
- Recruit and build awareness among students
- Adapting & developing curriculum in dynamic and evolving field
 - Hope to host an educational summit for during Digital Curation Centre's 6th International Conference in 2010, hosted by CIRSS at UIUC.





Questions & comments welcome

clpalmer@uiuc.edu

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