Purposeful Curation: Research and Education for a Future with Working Data

Carole L. Palmer, Allen H. Renear, Melissa H. Cragin
Center for Informatics Research in Science and Scholarship
Graduate School of Library and Information Science,
University of Illinois at Urbana-Champaign

Digital libraries are for users.

Is it true that "digital libraries are more akin to archives than they are to traditional libraries" (Ross, 2007)?

Like physical libraries, digital libraries exist to support the information needs of user communities, providing access to information in ways that add value and enhance use.

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

Libraries collect, curate, and then archive and preserve, with a purpose—the future use of scholarship, science, history, and heritage.

LIS is about information organization and access for user communities.

Is it true that "library science has not demonstrated that it has the theoretical foundations and knowledge base that are capable of providing the framework for handling digital entities and for underpinning digital libraries" (Ross, 2007)?

The theoretical foundations of Library & Information Science (LIS) go well beyond those of archival science to:

(i) user communities and their information behavior
(ii) data representation and retrieval, and
(iii) collection and service development and management.

Its theories and research are aimed at adding value to improve use (Taylor, 1986) and coordinating information in alignment with complex social structures (Shera, 1972).

LIS research and education advances curation for data use.

No one field has the range of theory and practice necessary for managing the entire lifecycle of digital content.

The contributions of many disciplines are needed for the development of high functioning digital libraries, repositories, and curation services.

The need for LIS contributions to the field is evident in results from our current research on scholarly and scientific data, digital collections, and our experiences with the Data Curation Educational Program (DCEP) masters and continuing education activities.

◆ Digital Humanities Centers Curation Project

interviews indicate community's core concerns

- Curation should be informed by our best understanding of how data will used by researchers and scholars, accommodating both current and emerging methods.
- Metadata is becoming data and needs to connect data across domain boundaries as well as object boundaries.
- Curation must scale in ways that accommodate changing formats and data models.
- Markup variation creates interoperability and transformation difficulties that require new tools and strategies.

◆ Environmental Data Management Needs Project

survey indicates local service priorities

Dealing with large amounts of data
- assistance with database design
- storage of data for collaborations

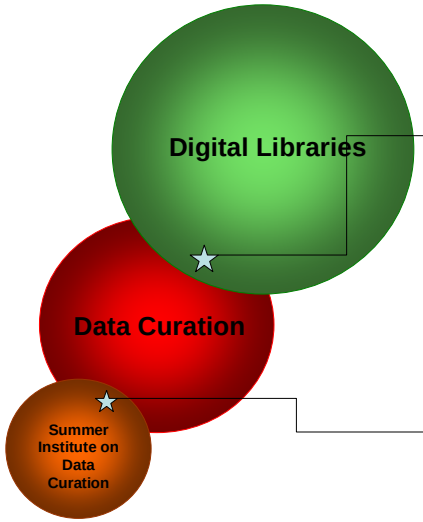
Migrating data & data conversion
- real time delivery of multiple data sources

Supporting disciplinary specialization
- data service technology for earth science
- cross-disciplinary geographic information system

◆ Digital Collection Aggregation Project

User testing shows problems representing scale and granularity in metadata and interface

- small window into large, diverse accumulation
- strengths for scholarly purposes not evident
- relationships among items and collections not exploited
- transformations and new composites not accommodated



Range of Shared Courses

Systems Analysis and Management
Digital Preservation
Metadata
Foundations of Data Curation
Digital Humanities
Information Retrieval
Digital Libraries
Document Modeling
Electronic Publishing
Information Interfaces
Information Modeling
Ontology Development
Representation & Organization of Info

Topics in Foundations Course

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
Scholarly Research Practices

Data Curation Educational Program (DCEP)

DCEP builds on existing LIS Digital Library curriculum and is focused on curation for scientific and scholarly research data. It includes a concentration for masters students in LIS and summer institutes for practicing academic librarians and other research data practitioners.

These new information professionals will build and maintain not only digital libraries and curated data sets, but also the associated indexing systems, metadata standards, ontologies, and retrieval systems.

◆ Curation Profiles Project

Interviews and case studies show variation in curation needs across sciences that must be accommodated in data repositories.

	Crystallography	Geobiology
Data Characteristics		
Type	1. "Raw data" - most information rich - long-term value for re-use 4. "CIF file" - most commonly shared data type	1. Large spreadsheet 2. "Reduced spreadsheet" - averaged values for multiple observations - most often requested by others
Format	1. Binary data – image 4. Crystallographic Information File (field-wide standard for numerical data)	2. Excel spreadsheet
Size	1. Image set is approx. 1Gb 4. > 500Kb	1. spreadsheet – under 1Mb
Intellectual Property	• Service model Ownership of the data is ambiguous, and requires negotiation before data "hand-off"	• Depends on source of funding (gov., private grants, industry) • Ownership of and rights to the data range from full to very limited
Will Share When?	Negotiated, often after 2 years - many journals require deposit of CIF files	Long-term "embargoes" sometimes required
Search and Retrieval	• Field-wide repositories • OAI-PMH tools becoming available for CIF files	• Difficult and ad hoc • Authors receive direct data requests

References

Ross, Seamus. (2007). Digital preservation, archival science and methodological foundations for digital libraries. Proceedings of the 11th European Conference on Digital Libraries (ECDL), Budapest (17 September 2007). Available: http://www.ecdl2007.org/Keynote_ECDL2007_SROSS.pdf. Accessed July 24, 2008.

Shera, Jesse H. (1971). The Complete Librarian and other essays. Cleveland, OH: The Press of Case Western Reserve University.

Shera, Jesse H. (1972). An epistemological foundation for library science. In J. H. Shera, The Foundations of Education for Librarianship (pp. 109-134). New York: Becker and Hayes.

Taylor, Robert S. (1986). Value-added processes in information systems. Norwood, N.J.: Ablex Publishing Corporation.

The projects described above have been funded primarily by the Institute of Museum & Library Services.

We wish to acknowledge the outstanding assistance provided by W. Aaron Collie in designing this poster, and Dr. Jerry McDonough for his original analysis of the overlap between the GSLIS Digital Library and Data Curation curriculum.