Data Curation Education

JCDL
Pittsburgh, June 20, 2008
Linda C. Smith
Overview

- Defining “data curation”
- The Data Curation Education Program
- Cooperating Institutions
- Advisory Committee and Other Contributors
- Building the Curriculum
- Program Summary
- Future Events
Data Curation is…

The active and on-going management of (research) data through its lifecycle of interest and usefulness to scholarship, science, and education.
Data Curation Activities

- Build collections (appraisal and selection)
- Enable data discovery and retrieval
- Provide for re-use over time
- Maintenance of data quality (integrity and authentication)
- Value-add services (metadata, adding links)
- Preservation (migrating data forward)
- Archiving
Range of Curation Services

- Repository for Data (and related supplementary materials)
- Deposition Support Services
- Discovery and Access Services
- Metadata Services (descriptive, preservation)
- “Linking” Services

- Appraisal & Selection
- Intellectual Property Rights Consultation
- Large-scale Storage (dark archive / back-up service)
- Migration / Emulation
- Archival “Packaging”
- Applications and Tools (visualization, social networking)
Emerging Curation Service Models

- Consultation – data management and curation services
- Institutional Repository as centralized data service approach
- Library Portal – services for distributed data collections
- Collaboration
  - Add library services directly into grant proposals
  - Library / laboratory develop tools together
- Research Services – provide “end-to-end” services for particular disciplines
- Mobilizing for e-Science – planning library involvement in cyberinfrastructure
Data Curation Education Program
Data Curation Education Program

- DC is a concentration within our MSLIS
- Developed on a foundation of research and teaching on scientific practice
  - biodiversity information problems
  - metadata and preservation
- Developed alongside our Biological Information Specialist program
- Synergy with Digital Library CAS and the MS in Biological Informatics (not MSLIS)
Data Curation Education Program

Goals

- Build educational capacity within GSLIS and at UIUC to train information professionals in digital data curation
  - New courses, case studies, continuing education
  - Expand understanding of the role of data curation in the production of research
- Build partnerships with institutions that will offer internship and practicum opportunities
- Develop best practices reports on the education of data curation professionals
- Offer the program on campus and via LEEP, our on-line option
DCEP Priorities

- Curriculum development
- Recruitment
- Development of field work opportunities
- Needs assessments
- Dissemination of information
Cooperating Institutions

Advisory Committee

- MIT Library
- Missouri Botanical Garden
- Smithsonian Institution
- US Geological Survey
- Field Museum
- ICPSR

Other Contributors

- Biomedical Informatics Research Network (UCSD)
- Marine Biological Laboratory
- Johns Hopkins U. Libraries
- U.S. Army Strategic Environmental Management Program (SEMP)
- National Library of Medicine
Advisory Committee Contributions

- Identify data management and curation problems
- Facilitate collection of best practices
- Develop case studies for use across the curriculum
- Facilitate internships
- Cultivate new partners and new collaborative research
Skill Sets Needed

- Knowledge of scholarly communication processes and how research works
- Domain knowledge (or access to it!)
- Ability to talk to domain experts, programmers, and technologists
- Ability to track emerging technology
- Knowledge of metadata (incl. disciplinary standards)
- Understanding variation in standards
Course areas needed

- Technical, Social (practice, organizational), Policy
  - Data-centric
    - Risk assessment, Preservation and Sustainability, Standards and Metadata, Intellectual Property
  - Scientific research process
    - Scholarly communication
    - Lifecycles
  - Domain-oriented courses
    - Biodiversity and EcoInformatics, Ontologies in Natural Science
  - Other
    - Databases, Project management
DCEP Curriculum

Required of All Master's Students
- LIS501 Information Organization and Access
- LIS502 (2 hrs only) Libraries, Information and Society

Required for the DC Concentration
- LIS590DC Foundations of Data Curation
- LIS590PD Digital Preservation
- LIS453 Systems Analysis and Management
- Field Experience Seminar (Req’d if taking a practicum, 2 hours)
DCEP courses, cont’d

DCEP List of Recommended Electives
(Students required to take two, we recommend four)

- LIS452 Foundations of Information Processing in LIS
- LIS590BDI Biodiversity and Ecoinformatics
- LIS590DI Digital Libraries: Research and Practice
- LIS590DM Document Modeling
- LIS590IM Information Modeling
- LIS590MD Metadata in Theory and Practice
- LIS590OD Ontology Development
- LIS590RO Representing and Organizing Information Resources
Core course content

Foundations of Data Curation

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

Assignments on:
- Analysis of Data Management Plans
- Discipline-based data curation needs assessment

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
DCEP Status

- First cohort (2007) – 6 students
- Second cohort (2008) – approx. 10 students
- Internship at Johns Hopkins University Libraries (2008)
Demand and related challenges

- More internship openings than students
- Recruit and build awareness among students
- Data curation and BIS students proving to be attractive to research centers and library operations
- Expressed need for continuing education
- Adapting & developing curriculum in dynamic and evolving field
- Hope to host an educational summit during the 6th Digital Curation Conference in 2010, to be held at UIUC.
Summer Institute on Data Curation

First Summer Institute June 2-5, 2008

- **30 participants** (academic librarians and administrators, library IT staff, LIS faculty, bioinformatics student, CISTI staff)
- **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 and authenticity
  - Appraisal and selection
  - Preparation for ingest
  - Digital preservation standards
  - Day-to-day preservation work
  - Repository architectures
Project Deliverables and Future Events

- Adding new contributor organizations
- Curriculum materials to be available via IDEALS, the UIUC institutional repository
  - syllabi and case studies
- Summer Institute on Data Curation
  - Providing infrastructure and strategy for on-going collaboration and communication of participants
- New IMLS grant award: “Extending Data Curation to the Humanities”
- GSLIS to host the 6th Digital Curation Conference in 2010
Data Curation Education Program

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

P. Bryan Heidorn, PI
Carole L. Palmer, Melissa H. Cragin, Linda C. Smith

{pheidorn, clpalmer, cragin, lcsmith @uiuc.edu}

Data Curation Educational Program (DCEP)
IMLS – Laura Bush 21st Century Librarian Program
RE-05-06-0036-06 (Heidorn, PI)

http://www.lis.uiuc.edu/programs/ms/data_curation.html