Data Curation Education and Biological Information Specialists

DigCCurr 2007
Chapel Hill, April 20, 2007
Overview

The Data Curation Education Program

- Integration of research and practice
- Related GSLIS research
- Cooperating Institutions
- Building the curriculum
- Advisory Committee Contributions
- Program summary
- Future Events
Data Curation Education Program

- Concentration in MSLIS
- Synergy with Digital Library CAS
- Synergy with Master of Science in Biological Informatics (not MSLIS)
- Data and Digital
Integration of research and practice

- Identify information problems and collect **best practices** from our partners to provide a broad understanding of information and data techniques, issues, and needs.

- Place students in **internships** with our partners at biological science institutions to gain real-world biological research experience.

- **Cultivate new partners** and new collaborative research.
Related GSLIS Projects:

Metadata and Preservation

Automated metadata extraction and inference
• Automatic Museum Label Metadata Extraction
  Heidorn, NSF DBI-9982849, NSF, DBI-0345387
• Georeferencing Museum Specimen Sources
  Heidorn, Moore Foundation 2005-2929-00

Metadata and Collections
• Digital Collections and Content
  Cole, IMLS National Leadership Grant LG-02-02-0281

Preservation
• EchoDep (NDIPP)
  Unsworth, Sandore, Library of Congress
Related GSLIS Projects: Data Centric

Modeling and computational neuroscience
  • Information and Discovery in Neuroscience
    Palmer, NSF IIS-0222848

Terminology, schema, and ontology development
  • Plant Description Standards
    Heidorn, IMLS NR-00-01-0017-01

Collaborative data collection
  • BioDiversity Survey Collaboration and Verification
    Heidorn and Palmer, NSF ITR/BDI-0113918
Cooperating Institutions

- Biomedical Informatics Research Network (UCSD)
- Marine Biological Laboratory
- Smithsonian Institution
- American Museum of Natural History
- Missouri Botanical Garden
- U.S. Army Strategic Environmental Management Program (SEMP)
- MIT Library
### New and proposed courses

<table>
<thead>
<tr>
<th>Course Title</th>
<th>Instructor</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Curation</td>
<td>Cragin</td>
<td>Fall 2007</td>
</tr>
<tr>
<td>Ontologies in the Natural Sciences</td>
<td>Renear</td>
<td></td>
</tr>
<tr>
<td>Information Transfer and Collaboration in Science</td>
<td>Palmer</td>
<td></td>
</tr>
<tr>
<td>Scientific Data and Procedure Standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discovery Informatics and Data Mining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Preservation</td>
<td>McDonough</td>
<td>Spring 2008</td>
</tr>
<tr>
<td>EcoInformatics / Biodiversity Informatics</td>
<td>Heidorn</td>
<td></td>
</tr>
<tr>
<td>Metadata in Theory and Practice</td>
<td>McDonough</td>
<td></td>
</tr>
<tr>
<td>Scientific Classification and Vocabulary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field Experience Seminar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(to accompany practicum)
Advisory Committee Contributions

- Identify information problems and collect **best practices** from our partners to provide a broad understanding of information and data techniques, issues, and needs.

- Develop **case studies** for use across the curriculum.

- Place students in **internships** with our partners at biological science institutions to gain real-world biological research experience.

- **Cultivate new partners** and new collaborative research.
Emergent Themes
Initial Advisory Committee Meeting

- Data-centric
- Policy and Sustainability
- Researchers and Research
- Information / Management

The items on the following slides include skills grouped according to their technical or social basis.
Data-centric

Technical
- File formats
- Metadata
- Ontologies
- Standards

Social
- Data practices
Policy and Sustainability

Technical

- Analysis and implementation of ownership models and copyright
- Preservation strategies
- Risk assessment

Social

- Risk assessment
- Social, cultural and political aspects of intellectual property
Researchers and Research

Technical

- Workflows (and automation)
  - Not just preservation workflow.

Social

- Domain knowledge
- Social parameters of research production
- Research lifecycle
- Scholarly communication
Information Management

Technical

- Use of applications across platforms (PC/Mac/Linux)
- Basic data file care: storage, backup
- Database management systems
- Systems evaluation

Social

- Recruit personnel
- Facilitate collaboration
- Change management
DCEP Guideposts

- Knowledge of practice
  - how things are done now (not always best)
  - best practices, application and generalization

- Knowledge of problems
  - data “problems” not yet resolved by current practice
  - research needed to enumerate and analyze these problems

- Knowledge of promise
  - awareness of technologies that could solve problems
  - societal information trends
    - Changing funder&scientist’s perspectives
Project Deliverables and Future Events

- Curriculum materials to be available via IDEALS, the UIUC institutional repository
  - syllabi
  - case studies

- Summer Institute in Data Curation
  - Summer, 2008 workshop for academic and research librarians

- National conference on the Practice, Problems and Promise of Data Curation in 2009/2010
New Directions

- Growing changing partnerships
- Live distance education
- Adjunct faculty and guest lectures
- Humanities
  - E.g. 2007 Digital Humanities Conference
- Research efforts in science and scholarship
- Science Communication Initiative
  - http://sci.lis.uiuc.edu/
Data Curation Education Program

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


{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