NCAR Library Update

Presentation to the Academic Affiliates
UCAR Annual Meeting

Mary Marlino
Jamaica Jones

October 6, 2010
Overview

- Building capacity for scientific data curation
- Debut of OpenSky!
Scientific Data Curation

• Data deluge
  • 80% of all research data produced by 3 disciplines: high energy physics, atmospheric science, astronomy*

• Future workforce issues
  • Critical shortage of data curation professionals in geosciences
  • Need to develop capacity in libraries and academic institutions
  • Diffusing data skills into the research community essential

*Harnessing the Power of Digital Data for Science and Society, 2009 report of The Interagency Working Group on Digital Data of the National Science and Technology Council’s Committee on Science
New Roles for Libraries

- Libraries, information scientists, and library educators now aligning with demands of the data deluge
- Supporting increased transparency, linking data to scholarship
- Supporting NSF-mandated research data plans
- Library schools creating new education programs
- Data providers and users are recognizing that traditional library skills can be brought to bear on issues of data curation, preservation, access
- Result: maximizing the accessibility, utility, preservation, and re-use of scientific data
National Calls

“Ongoing attention must be paid to the education of the professionals who will support, deploy, develop, and design current and emerging cyberinfrastructure.

For example, the increased emphasis on ‘data rich’ scientific inquiry has revealed serious needs for ‘digital data management’ or data curation professionals. Such careers may involve the development of new, hybrid degree programs that marry the study of library science with a scientific discipline.”
Library Collaborations to Build Capacity

- Institute of Library and Museum Sciences (IMLS) award
  - Partnership: Illinois, Tennessee, UCAR
  - Building leadership in data curation: next gen librarians and information scientists
    - Meeting the NSF challenge: *Growing the Geoscience workforce of the future*
  - Full funding for 4 PhDs, 4 Master’s students for 2-4 years
PhD Program at Illinois

- Specialization in Data Curation, in partnership with NCAR

- Emphasis: information science research in the curation of atmospheric and geosciences data

- Awarding funding for 4 years of tuition and stipend
  - Two years in research position at the Center for Informatics Research in Science
  - One year graduate research fellowship at NCAR working with science and data mentors
  - One-year dissertation fellowship
## Academic and research experience

<table>
<thead>
<tr>
<th>Selected Courses</th>
<th>Selected Research Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Problems in Data Curation</td>
<td>Cross-disciplinary data sharing and reuse</td>
</tr>
<tr>
<td>Ontologies in the Sciences</td>
<td>Data units, identity, and provenance</td>
</tr>
<tr>
<td>Information Transfer &amp; Collaboration in Science</td>
<td>Collection and integration of data for scientific problem-solving</td>
</tr>
<tr>
<td>Metadata in Theory &amp; Practices</td>
<td>Data and literature dynamics</td>
</tr>
<tr>
<td>Foundations of Data Curation</td>
<td>Culture and practice of data communities</td>
</tr>
</tbody>
</table>

Academics integrated with extensive experience with information science research teams and data center experts.

Details at [http://www.lis.illinois.edu/academics/programs/phd](http://www.lis.illinois.edu/academics/programs/phd)

Application deadline: 1/5/2011
: Master’s degree program at Tennessee

- Specialization in science data curation, in partnership with NCAR Library
- Emphasis: information science practice and education in the curation of science data and information.
- Funding for 2 years of tuition and stipend
  - Summer experience on site at NCAR working with science and data mentors
  - Participate in research projects at the Center for Information and Communication Studies working with research team and practitioners
  - Opportunity to work with scientists from Oak Ridge National Lab and related organizations
### Selected Courses

<table>
<thead>
<tr>
<th>Selected Courses</th>
<th>Practice and research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Problems in Data Curation</td>
<td>Cross-disciplinary data sharing and reuse</td>
</tr>
<tr>
<td>Information Organization and Representation</td>
<td>Data management practices of scientists</td>
</tr>
<tr>
<td>Sources and Services in Science &amp; Engineering</td>
<td>Participation in research teams</td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>Shadowing scientists in laboratory environments</td>
</tr>
<tr>
<td>Practicum and Research Participation</td>
<td>Culture and practice of data communities</td>
</tr>
</tbody>
</table>

Academics integrated with extensive experience with information science research teams and data center experts in the science domain.

Details at: [http://www.sis.utk.edu](http://www.sis.utk.edu)

Rolling admissions
Introducing OpenSky!
What is OpenSky?

- The institutional repository of UCAR, NCAR and UCP
  - Published, peer-reviewed articles
  - NCAR Technical Notes
  - UCAR-authored manuscripts and monographs
  - Proceedings, presentations, and posters
  - Educational materials
- Realization of UCAR’s Open Access policy
What is the UCAR Open Access policy?

- Passed September 2009
  - “…each employee will provide an electronic copy of the final version of his/her Scholarly Work in an appropriate format at no charge to the NCAR Library.”
  - “The NCAR Library will take reasonable measures to make the Scholarly Works available to the public, including through an open-access repository.”
What does OpenSky do?

- OpenSky:
  - supports the UCAR community by providing free, and open access to UCAR research
  - increases the visibility and impact of that research
  - provides information regarding authors’ rights
  - supports internal reporting and metrics requirements
What will you find in OpenSky?

- Opening Day Collections
  - Roughly 2,300 full-text resources
  - ~200 peer-reviewed journal articles
    - AMS – 6 mo. Embargo, 100 most-cited
    - AGU – 6 mo. embargo
    - Springer – No embargo
  - ~500 NCAR-authored Technical Notes
  - 15 years of SOARS protégés’ papers
- Archival materials
  - Warren M. Washington’s personal papers
  - Documentation of the construction of the Mesa Lab
- ~50 years of Staff Notes
What will you not find in OpenSky?

- Works that are under embargo by the publisher
- Datasets
- (Most) works published before November, 2009
What does OpenSky mean to you?

- Timely and free access to UCAR scholarship
Demonstration
## Data Curation Doctoral Program

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

<table>
<thead>
<tr>
<th>Doctoral Program Phases</th>
<th>DCERC Program Components</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase 1 (yrs 1 &amp; 2)</strong></td>
<td><strong>Coursework (48 credit hours)</strong></td>
</tr>
<tr>
<td></td>
<td>Research Assistantship (GSLIS) (data curation-related projects at GSLIS)</td>
</tr>
<tr>
<td></td>
<td>- Doctoral seminar: Research Problems in Data Curation</td>
</tr>
<tr>
<td><strong>Phase 2 (yr 3)</strong></td>
<td><strong>Field Exam</strong> (Fall semester)</td>
</tr>
<tr>
<td></td>
<td>- written and oral exam</td>
</tr>
<tr>
<td></td>
<td><strong>Research presentation</strong> (Year 2 or 3)</td>
</tr>
<tr>
<td></td>
<td>Year-in-Residence at NCAR (funded)</td>
</tr>
<tr>
<td></td>
<td>- Doctoral seminar: Information Transfer &amp; Collaboration in science</td>
</tr>
<tr>
<td><strong>Phase 3 (yrs 3 &amp; 4)</strong></td>
<td><strong>Dissertation</strong></td>
</tr>
<tr>
<td></td>
<td>- Defense of a dissertation proposal (Year 3, Spring sem. preferred, or Year 4, early Fall)</td>
</tr>
<tr>
<td></td>
<td>- Final defense and submission of dissertation</td>
</tr>
<tr>
<td></td>
<td>Dissertation fellowship year</td>
</tr>
</tbody>
</table>

http://www.lis.illinois.edu/academics/pro
<table>
<thead>
<tr>
<th>Year 1: Fall and Spring</th>
<th><strong>Education</strong></th>
<th><strong>Experiences</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coursework (18 credit hours) core and electives including a focus on science information and data curation</td>
<td>Research Assistantship (SIS) (science data related projects)</td>
</tr>
<tr>
<td>Year 1: Summer</td>
<td>Coursework (3 credit hours)</td>
<td>Summer-in-Residence at NCAR (funded)</td>
</tr>
<tr>
<td>Year 2: Fall and Spring</td>
<td>Coursework (18 credit hours)</td>
<td>Work with mentors at ORNL and related science agencies</td>
</tr>
<tr>
<td>Year 2: Summer</td>
<td>Coursework (3 credit hours) And comprehensive examination for graduation</td>
<td>Summer-in-Residence at NCAR (funded)</td>
</tr>
</tbody>
</table>

http://www.sis.utk.edu/welcome
Scientific Data Curation & Libraries

- **Data deluge**
  - 80% of all research data produced by 3 disciplines: high energy physics, atmospheric science, astronomy

- **Future workforce issues**
  - Critical shortage of data curation professionals in geosciences
  - Need to develop capacity in libraries and academic institutions
  - Diffusing data skills into the research community essential

- **New roles for libraries**
  - Libraries, information scientists, and library educators now aligning with demands of the data deluge
  - Responding to increased role of libraries in data-intensive research
    - Supporting increased transparency, linking data to scholarship
  - Data providers and users are recognizing that traditional library skills can be brought to bear to issues of data curation, preservation, access
  - Result: maximizing the accessibility, utility, preservation, and re-use of scientific data