



Data Conservancy: A Library- Based Data Cyberinfrastructure Paradigm

Melissa Cragin

Center for Informatics Research in Science and Scholarship
Graduate School of Library and Information Science
University of Illinois

Salmon Data Access Working Group – CERF Informatics Meeting

November 4, 2009





Data Conservancy



- Award under the NSF DataNet program
- Library–led organization

G. Sayeed Choudhury, PI, Johns Hopkins University, Sheridan Libraries

- Addressing “grand challenges”
(scientific use cases; cyberinfrastructure)
- Facilitating new science
 - access to “small science” data
 - cross-disciplinary research





NSF DataNet Partners: Three Primary Goals

- Achieve long-term preservation and access capability in an environment of rapid technology advances.
- Create systems and services that are economically and technologically sustainable.
- Empower science-driven information integration capability on the foundation of a reliable data preservation network.



Data Curation

... is the active and on-going management of (research) data through its lifecycle of interest and usefulness to scholarship, science, and education.

The Data Conservancy embraces a shared vision: data curation is **not an end, but rather a means** to collect, organize, validate, and preserve data to address grand research challenges that face society.





Core components of the DC

- data model for observations
 - identity; date/time stamp; location; discipline-specific attributes
- modular architecture
- reference implementation using Fedora
- access to data stored / preserved elsewhere
- semantic technologies
 - OAI-ORE (Open Archives Initiative Object Reuse and Exchange)
- sustainability





Partner institutions

- Johns Hopkins University (Lead institution)
- Cornell University
- Encyclopedia of Life (Marine Biological Laboratory)
- Fedora Commons
- National Center for Atmospheric Research
- National Snow and Ice Data Center
- Portico
- Tessella, Inc.
- University of California Los Angeles
- University of Illinois at Urbana-Champaign





Information Science Research

	Astronomy	Biodiversity	Earth Science	Social Science	
UCAR	Task-based design and usability testing ⇒ User cases, data requirements, system recommendations				
UCLA	Ethnography, virtual ethnography, oral histories ⇒ Use cases, data requirements	Interviews, Surveys, Worksheets, Content analysis ⇒ Curation requirements, taxonomy, metadata/provenance framework			UIUC CIRSS





UIUC Research

○ Data Practices

- comparative analysis of disciplinary differences in data practices
- taxonomies on data, practice, & curation

○ Data Concepts

- development of formal terminology and identity conditions
 - collections, data sets, versions, and data items
- rules that relate collection and data set metadata.

○ Automated Learning Group

- extend SEASR (software Environment for the Advancement of Scholarly Research) to support DC research and education activities





UIUC Education Activities

- Extend internship and practicum opportunities for students to DC research teams and partners
- Web-based clearinghouse and forum for data curation educators



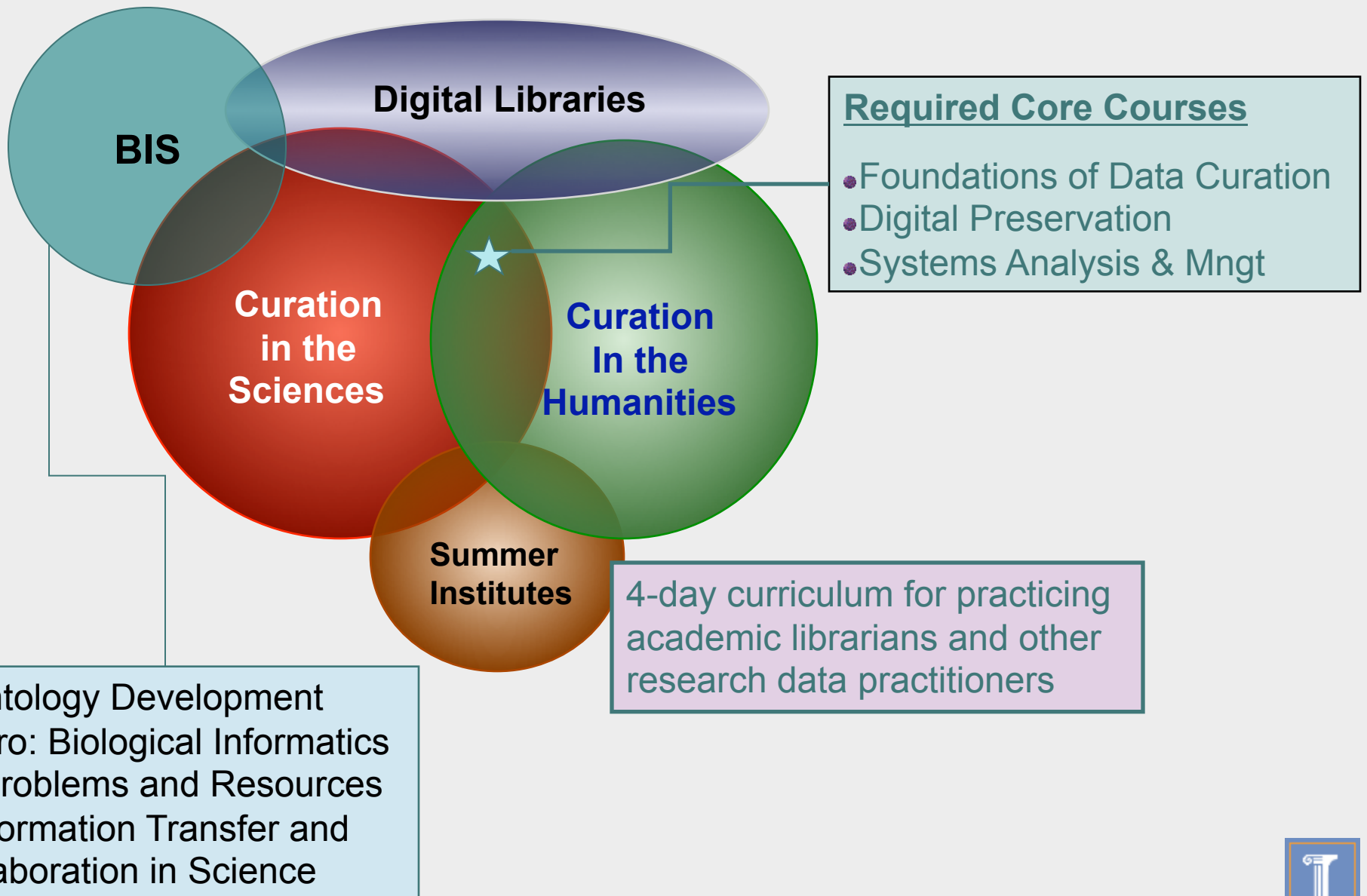


Data Curation Education Program

- Specialization in our Master of Science in LIS
- Field work opportunities
 - practica and internships
- Advisory Committee and Partnerships
- curation needs assessment
- Summer Institute on Data Curation for in-service information professionals



Building a workforce through professional education

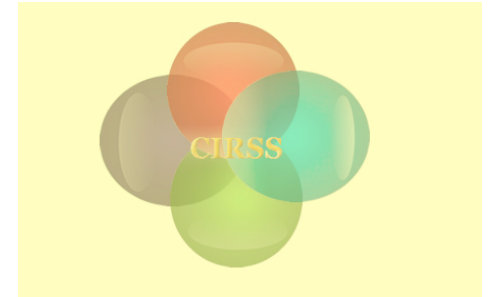
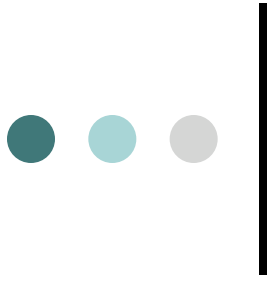




Data Conservancy: Implications for Scientific Research

- Libraries as part of a distributed network
 - data as collections, data as services
- Focus on facilitating new science
- Sustainability of data network
 - leverage partners' sustainability mechanisms
 - adaptable infrastructures
 - business requirements





Thank you

