Evaluating the Uses of Digital Libraries

DELOS Workshop on Evaluation of Digital Libraries

Padova, Italy, 4 October 2004

Christine L. Borgman

Professor & Presidential Chair in Information Studies

University of California, Los Angeles

&

Oxford Internet Institute

University of Oxford, UK

Outline of Talk

- DELOS WP 7 Goals
- Use and context for digital libraries
- Case studies
  - Alexandria Digital Earth Prototype (ADEPT)
  - Center for Embedded Networked Sensing (CENS)
- Measures and metrics
- Discussion and conclusions

Task 7.2: Evaluation Models and Methods

- Integrated research on DL evaluation
- Initial focus on specification of standard DL evaluation methods
- Starting with comparison and evaluation of existing evaluation methodologies
- $\rightarrow$ DL evaluation workshop

Digital Libraries

- *Systems* that support searching, use, creation of content
- *Institutions* with people, digital collections, and services
- *Repositories* of digital data and documents, as a component of cyberinfrastructure / e-science / e-social science (etc.)
  - Primary data: scientific data from sensors, labs, field work
  - Secondary sources: published articles, monographs, reports
  - Teaching resources: lectures, labs, exercises, exams, illustrative documents and images

Cyberinfrastructure / e-Science

- Link human expertise, data, information, computational models, sensor arrays, specialized facilities
- Create new pathways for research
- Create "cyberinfrastructure enabled knowledge communities"
- Create community-specific knowledge environments for research and education (Atkins, 2004)

“Knowledge Communities” and Digital Libraries

- What are the scope and boundaries of “knowledge communities”?
  - Disciplines?
  - Collaboratories?
  - Workgroups?
  - Epistemic cultures?
- What is the relationship between digital libraries and “knowledge communities”?
  - Cyberinfrastructure enables new forms of distributed collaboration
  - Data sources, shared repositories, are essential components of scientific collaboration”
  - Sharing of resources is economically efficient for researchers, institutions, funding agencies, and societies (David & Spence, 2003)

Primary and secondary resources

- Digital libraries of secondary resources
  - Published documents
  - Scholarly products
  - Record of research
  - Institutional role of libraries and librarians
- Digital libraries of primary sources
  - Raw data from research
    - Instrumented data collection (labs, sensor networks)
    - Field notes
  - Archival sources
    - Unique documents
    - Records of individuals and organizations

Secondary sources (scholarly literature)

- Community orientation of researchers
  - publications are "end product" of research
  - incentive and reward system is based on publication
  - researchers contribute to digital collections (via publication)
  - publications are shared within invisible college
- Individual orientation of students
  - searchers of digital collections, not contributors
  - reliant upon search mechanisms and bibliographic control
- Digital libraries are "boundary objects" between experts and novices in a scholarly domain

Primary sources (scientific data)-1

- Community orientation of researchers
  - Scientific databases are becoming “end product” of research in some fields
  - Practices for sharing scientific data are evolving along with development of cyberinfrastructure
  - Sharing practices may vary widely by research area
  - Establishing agreements for access to data, credit for publications, is fundamental to any collaborative project
- Providing context to interpret data
  - Scholarly publications *provide* context
  - Digital libraries *remove* context

Primary sources (scientific data)-2

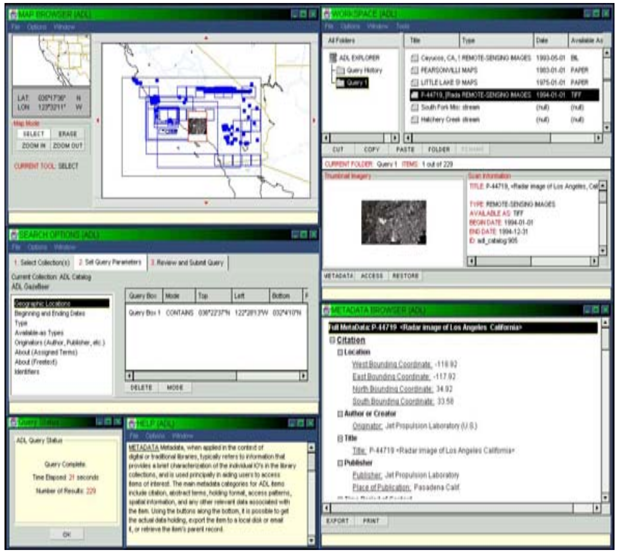
- Incentives to share data
  - Establish trust and reciprocity within a research group
  - Ability to mine large data sets, compare results
  - Ability to replicate experiments, studies
  - Requirement of some funding agencies
- Incentives *not* to share data
  - Rewards for publication, not for data management
  - Benefits of contributing data may accrue to other parties
  - Risks of others analyzing and publishing your data
  - Risks of misinterpretation of your data
  - Risks of losing control over data
  - Risks of loss of intellectual property

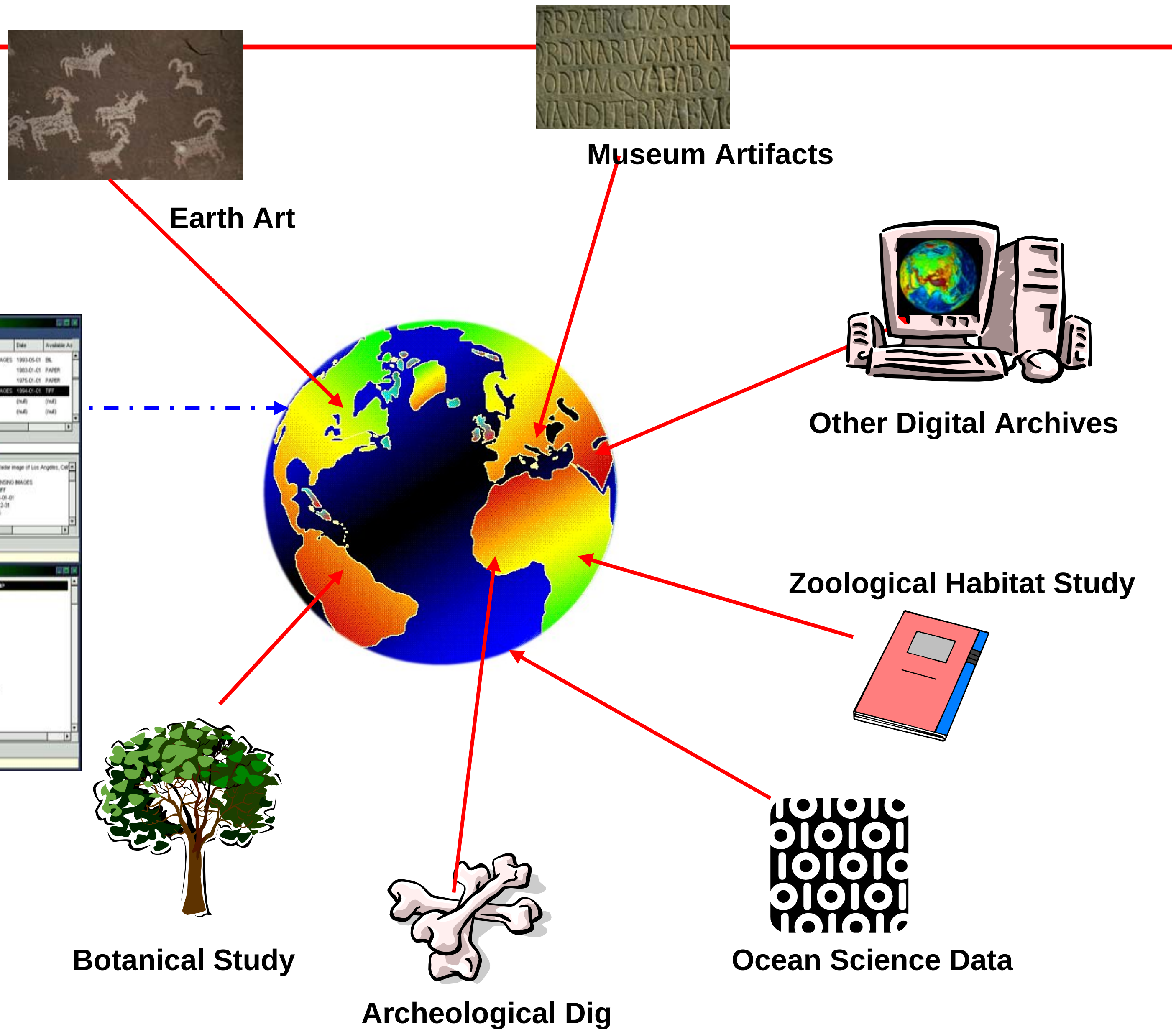
Evaluating primary source DLs in context

- Challenge: Design scientific digital libraries that will support research *and* teaching applications
- Goals:
  - Leverage investment in scientific data
  - Improve science instruction via inquiry learning
  - Provide services to use and share these data
  - Evaluate usefulness of digital libraries
- Case studies:
  - Alexandria Digital Earth Prototype (ADEPT)
  - Center for Embedded Networked Sensing (CENS)

Alexandria Digital Earth ProtoType

Alexandria DL of Distributed Spatial Information Objects





If *it* has a *latitude* and *longitude* then it can be in ADL

ADEPT Project: Geospatial digital libraries

Goals

- Add services to Alexandria Digital Library for teaching undergraduate courses in geography
- Facilitate inquiry learning by providing access to primary sources

User communities

- Faculty, as researchers
- Faculty, as teachers of undergraduate courses
- Undergraduate students

Activities to be supported

- Information searching and retrieval
- Composing lectures that incorporate text, concepts, and objects
- Constructing learning modules in which students can formulate and test hypotheses

Socio-technical studies and methods-1

1. Iterative design and classroom deployment of prototype virtual learning environments
   - Classroom observations, interviews with faculty, students, teaching assistants
   - Analysis of teaching materials (lectures, assignments, exams)
2. Faculty perspectives on the use of digital libraries for teaching geography
   - Interviews in faculty offices
3. Teaching assistant roles in the use of information technology for instruction
   - Interviews, observations in lab sessions
4. Faculty information seeking for research and teaching
   - Interviews in faculty offices

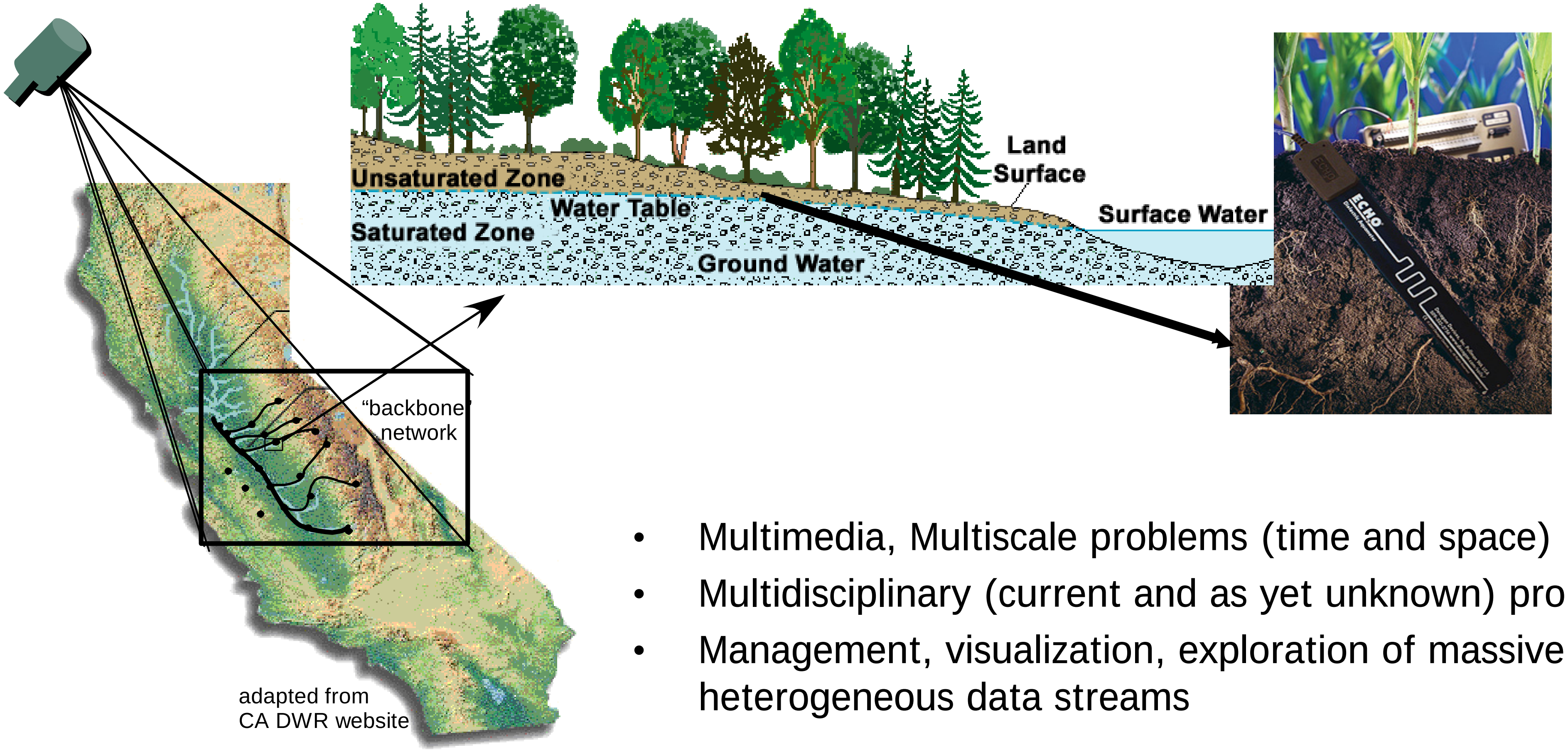
Socio-technical studies and methods-2

5. Student use of primary sources for inquiry learning
   - Interviews with students and faculty; analysis of student work
6. Adoption of digital libraries for undergraduate instruction
   - Assessment of take up rate for prototypes
7. Concept maps: How geography instructors organize teaching concepts
   - Classroom observations, videotaping, interviews
8. Metadata requirements for educational applications of geospatial digital libraries
   - Analysis of search queries, information seeking behavior, comparison to available metadata standards

Some ADEPT Results (1999-2004)

- Information seeking by geographers
  - Research: typical library use, online searching
  - Teaching: irregular, non-directed, often a by-product of research activities
- Information resources used by geographers
  - Research: varies by specialty; all want maps and images
  - Teaching: varies by course; all want maps and images
- Search queries of geographers
  - Research: concept, place (place name, latitude/longitude)
  - Teaching: concept, place, process (examples of erosion, population movements, etc.)
- Use of primary data in instruction
  - Preference for use of own research data
  - Tools to manage own research data would make DL teaching services more attractive

Data models for habitat monitoring and sensor networks



adapted from
CA DWR website

- Multimedia, Multiscale problems (time and space)
- Multidisciplinary (current and as yet unknown) problems
- Management, visualization, exploration of massive, heterogeneous data streams

Center for Embedded Networked Sensing: Education and Data Management Projects

- Goals
  - Make data from sensors useful for scientists on our research team and for other scientific communities
  - Make data from sensors useful for teaching high school science
  - Facilitate inquiry learning by providing access to scientific data by teachers and students
- User communities
  - Research scientists (habitat ecology, seismology)
  - High school science teachers (biology and physics)
  - High school students
- Activities to be supported
  - Scientific data management by scientists
  - Constructing learning modules in which students can formulate and test hypotheses
  - Experimental design and execution by “tasking” sensors for students

Methods and metrics

- Formative evaluation
  - Attending workgroup meetings of scientific teams
  - Analyzing work products of scientific teams (datasets, websites, publications)
  - Interviewing individual faculty
  - Visiting primary research site
    - Two-day research retreat at James Ecological Reserve, August 2004
  - Identification and assessment of available
    - Data repositories
    - Metadata standards and structures
  - Collaboration with ecology and seismic teams to assess repository requirements

Some CENS Results-1

- CENS has committed to sharing data; Center participates in NEON, NEESgrid, and related initiatives
- Maturity of data management practices varies widely by knowledge community
  - Seismic: Contributing data to community repository (IRIS) in standard format (SEED) for many years
  - Habitat ecology: Recent commitment to community repository (Morpho) in standard format (environmental metadata language); not yet implemented
  - Avian biology (localization of birdsongs): Sophisticated knowledge of data management issues, draws on practices from multiple disciplines
  - Education: Standards exist but high school teachers have little or no knowledge of them

Some CENS Results-2

- No metadata models exist that will address needs of all CENS scientific applications
  - Discipline / community specific standards needed
    - Environmental Metadata Language for biocomplexity data
    - SEED for Seismic data
  - Technology standards may bridge scientific communities
    - Sensor Markup Language to describe instruments
  - Geospatial coordinates required for most applications
    - Geospatial data standards exists for 2D points
    - Context descriptors also needed (distance from sea level, local distance from ground, above/below leaf, north/south side of tree)

METADATA FOR SENSOR DATA FOR HABITAT MONITORING			METADATA FOR EDUCATION MODULES FOR HABITAT MONITORING		
CENS Schema	**SensorML**	**EML 2.0**	**LOM**	**GEM**	**ADN**
CENS_Node.Node_Name Name of Node	***Sml:IdentifiedAs*** (2.2.2)				
CENS_Node.Node_Desc Description of Node	***AssetDescription : sml:description*** (2.2.12)				
CENS_Location.Location_ID Unique location ID	***CrsID*** (2.2.5)	***Eml-Coverage*** (2.4.4)			
CENS_Location.X_Pos (Position on X axis)	***HasCRS*** (2.2.5) ***ObjectState*** (3.3.6)	***Eml-Coverage-GeographicCoverage*** (2.4.4)			
CENS_Location.Time_Recorded Time location was captured		***Eml-Coverage-TemporalCoverage*** (2.4.4)			
CENS_Location.Time_Type_ID Refers to type of time of Time_Type ID table		***Eml-Coverage*** (2.4.4)			
			Educational-Typical Age Range (5.7)	***Audience-Age***	***Audience***
			Life Cycle-Contribute (2.3)	***Creator***	***Resource Creator***
			General-Coverage (1.6)	***Coverage-Spatial, Temporal***	***Coverage (spatial and temporal)***
			Life Cycle-Date (2.3.3) ***DateTime*** (8)	***Date***	***Creation date Accession date***
			General-Description (1.4)	***Description***	***Description***
			Educational (5)	***Pedagogy***	***Educational***

CENS Research Directions

- Infrastructure goals for CENS
  - Support scientists' requirements for collecting, managing, preserving, sharing data
  - Develop modular, extensible metadata architecture (XML-based)
  - Develop filtering tools to extract and visualize scientific data for educational applications
- Conduct behavioral studies of scientists, teachers, and students
  - How do they determine their data requirements?
  - What are their criteria for selecting, preserving data?
  - How do they use scientific data?
  - How do their uses evolve over time?
  - What are their incentives and disincentives to contribute data to repositories?

Some potential methods and metrics

- Goal: Sustainability of digital library
  - Transfer of tools between participants
  - Adoption of standards
  - Evidence of scalability
- Goal: Usefulness to a community
  - Evidence of contributions to shared repository
  - Evidence of adoption, take up, use in practice
  - Evidence of using, enhancing available tools
  - Evidence of re-use of contributed content

Discussion and Conclusions

- Digital libraries may have a wide range of users and of uses
- Users and uses interact in complex ways
- Cyberinfrastructure / e-science may enable new forms of collaboration and use of digital resources
  - These are claims to be assessed; not results
  - Research on the interaction between uses and users of CI are needed
  - Research is all about context, and DLs tend to remove context
  - Incentives and disincentives to use DLs exist
- Evaluation of use
  - Real measure is whether the DL is used

Acknowledgements

- ADEPT is funded by National Science Foundation grant no. IIS-9817432, Terence R. Smith, University of California, Santa Barbara, Principal Investigator.
  - ADEPT collaborators: Gregory Leazer, Anne Gilliland, Richard Mayer
- CENS is funded by National Science Foundation Cooperative Agreement #CCR-0120778, Deborah L. Estrin, UCLA, Principal Investigator.
  - CENS collaborators: Jonathan Furner, William Sandoval, Noel Enyedy, Michael Hamilton