

The Data Conservancy: Curating Data for Re-use

Mary Marlino

NCAR Library

National Center for Atmospheric Research

[CI Days: Cyberinfrastructure 2010 in the Rockies](#)

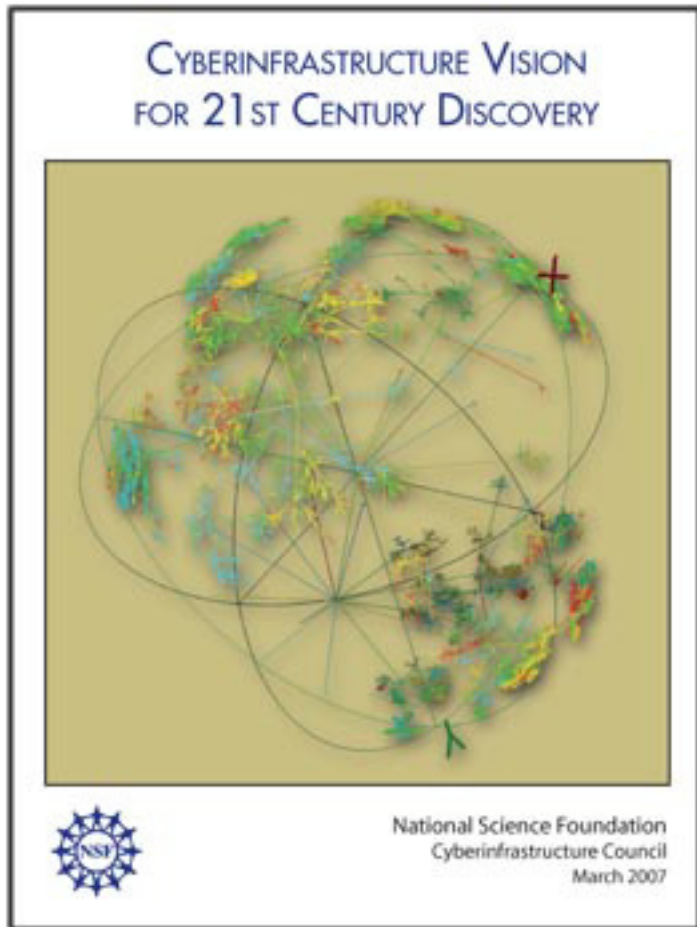
Data Curation and Digital Repositories Panel

August 13, 2010

Overview

- NSF DataNet program and goals
- Data Conservancy partnership and goals
- Implications for Libraries
- How to get involved

Sustainable Digital Data Preservation and Access Network Partners (DataNet)



Vision:

“...science and engineering digital data are routinely deposited in well-documented form, are regularly and easily consulted and analyzed by specialists and non-specialists alike, are openly accessible while suitably protected, and are reliably preserved.”

Understanding Infrastructure: Dynamics, Tensions, and Design



Report of a Workshop on “History & Theory of Infrastructure:
Lessons for New Scientific Cyberinfrastructures”

Paul N. Edwards
Steven J. Jackson
Geoffrey C. Bowker
Cory P. Knobel

January 2007



...not a rigid road map but principles of navigation. There is no one way to design cyberinfrastructure, but there are tools we can teach the designers to help them appreciate the true size of the solution space – which is often much larger than they may think, if they are tied into technical fixes for all problems.

NSF DataNet Program Goals

- Provide systematic, long-term preservation, access and analysis capabilities in an environment of rapid technology advances
- Engage at the frontiers of science and engineering research and education
- Serve as part of an interoperable data network spanning national and international boundaries



NSF Office of Cyberinfrastructure

DataNet Partner Requirements

- Combine expertise in library and archival sciences; computer, computational and information sciences, cyberinfrastructure; domain sciences and engineering
- Develop models for economic and technological sustainability over multiple decades
- Work cooperatively to create a functional data network with revolutionary new capabilities for access, use, and integration

The Data Conservancy (DC)

- DC is one of first two awards through the DataNet program
- Led by Sheridan Libraries at Johns Hopkins University
- DataONE: Observation Network for Earth, led by University of New Mexico Libraries
- Next round of DataNet will add up to three more partners into the network

Data Conservancy Partnership

DC is a network of domain scientists, information and computer science researchers, enterprise experts, librarians, and engineers

PI: Sayeed Choudhury—Sheridan Libraries, Johns Hopkins University

Co-PIs and Partners:

Carl Lagoze—Cornell University

Mary Marlino—National Center for Atmospheric Research (NCAR/ UCAR)

Carole Palmer—CIRSS, GSLIS, University of Illinois at U-C

Paddy Patterson—Marine Biological Laboratory

University of California Los Angeles
National Snow and Ice Data Center
DuraSpace/Fedora Commons

Tessella, Inc.
Portico



Other DC Partners

Australian National Data Service

Australian National University

British Library

Digital Curation Centre

Microsoft Research

Monash University

Nature Publishing Group

Optical Society of America

Sakai Foundation

Space Telescope Science Institute

SPARC

Sun Microsystems (Data Curation
Center of Excellence)

University of Queensland

Zoom Intelligence

Data Conservancy Goal

- Support new forms of inquiry and learning through the creation, implementation, and sustained management of an integrated and comprehensive data curation strategy
- DC 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

DC Objectives

- Infrastructure research and development
 - Technical requirements
- Information science and computer science research
 - Scientific or user requirements
- Broader impacts
 - Educational requirements
- Sustainability
 - Business requirements

Understanding Scientific and User Needs

Multi-site user research methods are a blend of:

- Case study and domain comparisons
- Depth and breadth
- Local and global

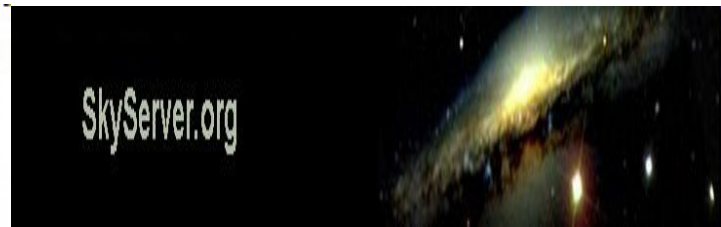
	Astronomy	Life Sciences	Earth Sciences	Social Sciences	
UCAR		Task-based design and usability testing ⇒ Use cases, data requirements, system recommendations			UCAR
UCLA	Ethnography, virtual ethnography, oral histories ⇒ Use cases, data requirements	Interviews, Surveys, Worksheets, Content analysis ⇒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

Research Questions

- **Data practices:** What are the data management, curation, and sharing practices?
- **Networks:** Who uses what data when, with whom, and why?
- **Curation:** What data are most important to curate, how, and for whom?

Astronomy as an exemplar scientific community

- Achieved notable success in community data standards, practices, documentation, and associated services for research and learning
- DC initial goal - ingest astronomy data into preservation archive, connect data to existing services used by astronomers
- Demonstrate utility of hosting data in environment that supports existing scientific capabilities in a sustainable manner

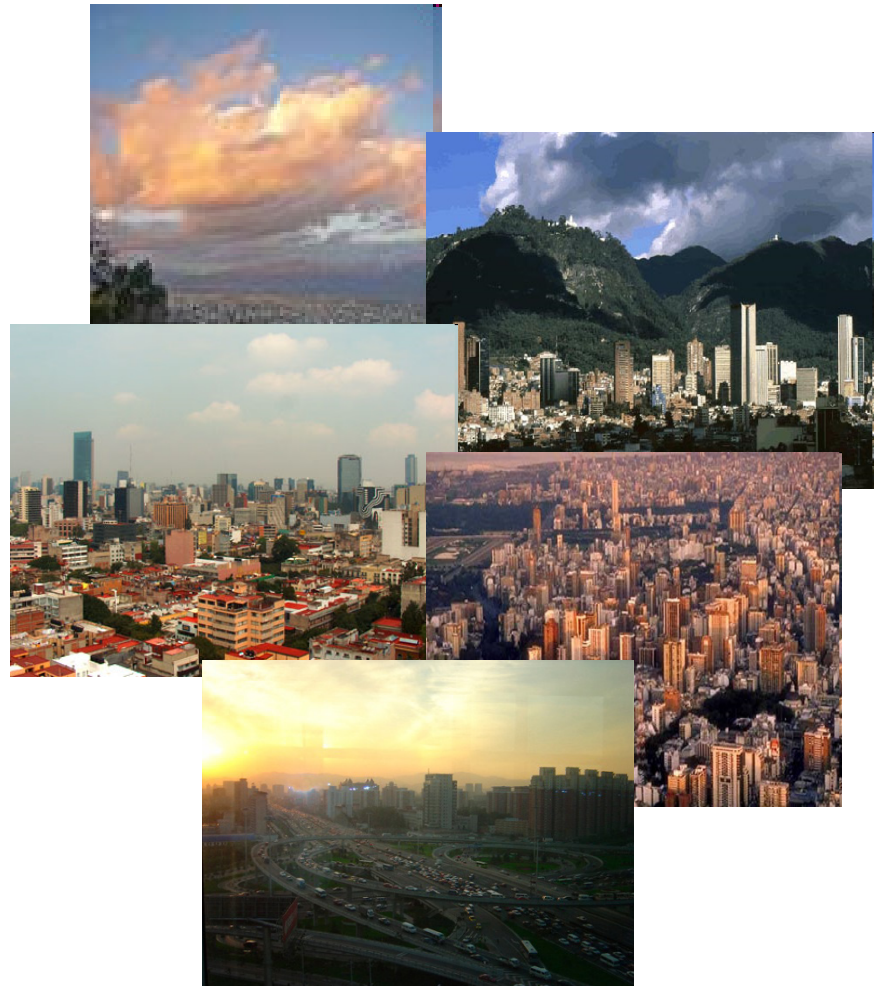




Social/Climate Science Research

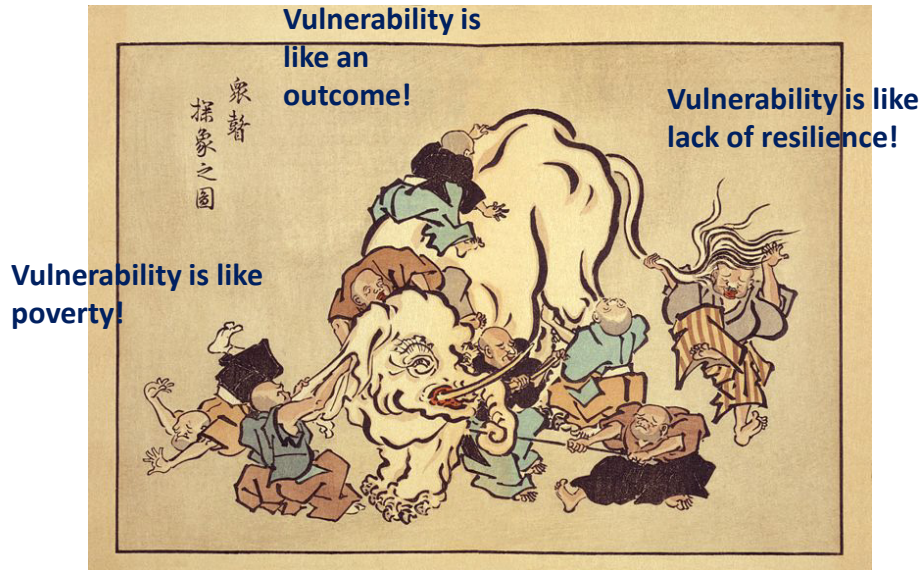
Assessing patterns of vulnerability/
adaptive capacity to
climate change
across urban areas

- emphasis on complex and heterogeneous data produced by different disciplinary domains



Urban Vulnerability

Paradox of the “Blind monks and the elephant”



An 1888 ukivo-e print by Hanabusa Itchō

- Complexity of urban vulnerability driven by
 - Array of hazards
 - Different units of analysis (affected sectors)
 - Specificities of
 - urban development
 - socio-environmental change
 - governance across cities

Broader Impacts and Educational Outreach

- Ensuring the wider community is involved with and will benefit from the infrastructure being developed
- Data curation outreach and education
 - Professional degree programs, in-service professional development, certification and institutes at Library/Information schools
 - Mentoring and “boot camps”
 - Field work practica and internships
 - Extending programs to educate more diverse set of students
 - Fellowships for students from traditionally underserved populations
- Communications on DC outcomes to university, scientific, and citizen stakeholders

Implications for Libraries

- Libraries as part of a distributed network
- Data as collections
- Data as services
- Librarians as data scientists/managers
- New requirements for Data Management Plans

“Data centers are the new library stacks”

– Winston Tabb (JHU Dean of Libraries)

How to Get Involved

- Be aware of new roles and opportunities for library professionals
- Investigate curricula and education programs in data curation such as Data Curation Education Program (DCEP) at the iSchool at Illinois
- Attend workshops and other professional development activities
 - <http://www.dcc.ac.uk/events/conferences/6th-international-digital-curation-conference>
- Stay informed of Data Conservancy and other DataNet project developments

Acknowledgements



Office of Cyberinfrastructure DataNet Award #0830976

Data Conservancy Partnership

Sayed Choudhury, Johns Hopkins University

Christine Borgman, UCLA

Carole Palmer and Melissa Cragin, Illinois

Contact marlino@ucar.edu for more information



NCAR



PORTICO





Thank you !