Cross-disciplinary Reuse

Carole L. Palmer
Director and Professor
Center for Informatics Research in Science & Scholarship
Graduate School of Library & Information Science
University of Illinois

After September 1, 2014
Information School
University of Washington

cyberinfrastructure

phases & relationships

reusable data resources

dimensions & dependencies

cross-disciplinary access and use

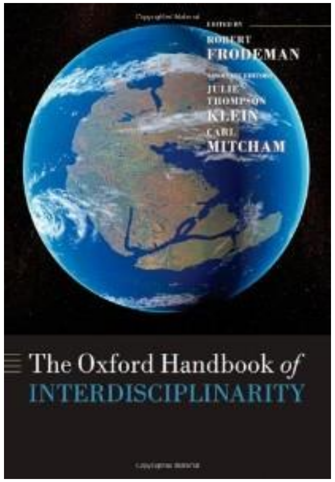
impact & value

New opportunity to get serious about cross-disciplinary cyberinfrastructure

Disciplinary strengths needed for multi-disciplinary research

(Seidel, 2011 e-mail shared with UIUC taskforce)

Greatest challenges for information systems is not ability to move across disciplinary boundaries but in maintaining the increasingly long and mutable intellectual pathsto our disciplinary past.



(Palmer, 2010, Oxford Handbook of Interdisciplinarity)

Studies of long-tail Earth & life sciences

Curation Profiles Project
2007-2009

Anthropology
Plant sciences
Kinesiology
Speech and Hearing
Earth and Atmospheric

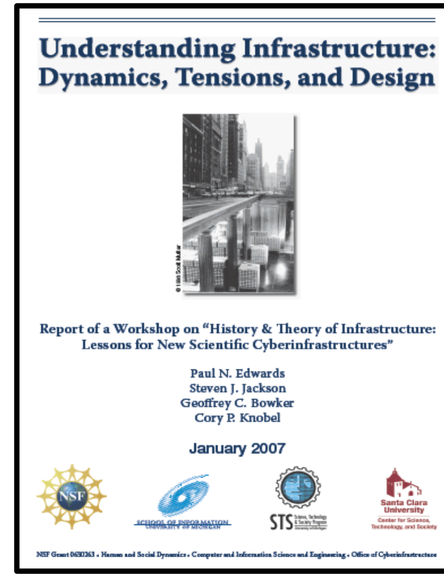
Oceanography
Climate science - modern
Climate science - paleo
Soil ecology
Volcanology
Stratigraphy
Mineralogy
Microbiology
Sensor network science
Environmental engineering
Photonics

Infrastructure – phases & relationships

Systems → Networks → Inter-networks

from <u>homogeneous, centralized</u>, local
to <u>heterogeneous, distributed</u>, coordinated

- technology growth and transfer
- consolidation
- gateways for interoperation



(Edwards, et al., 2007)

<u>RDA</u> explicitly addressing

technical / social – local / global

ground-up design; "<u>make-or-break" phase</u> (Parsons & Berman, 2013)

- Early choices constrain options

Infrastructure – reverse salients

- middleware specifications
- diverse data formats
- collecting & standardizing metadata

preservation

incentives

scaling

sustainability

Dimensions

disciplinary culture ?

human infrastructure

Paradigm dimension

Four

Fou~~r~~th paradigm **s**

- Data intensive science not "sweeping away the old reality" for a "paradigm shift in the Kuhnian sense."

- Continuum of modern scientific method

  includes and extends previous paradigms of

  empiricism, theory, and computation. (Wilbanks, 2009)

- Weak & strong coupling, dependencies

  (scoping cyberinfrastructure for combustion research)

Dimensions within and across disciplines

- **Open / closed** evidentiary cultures
  within gravitational wave research
  ("The Meaning of Data "Collins, 1998)

  - What qualifies as a publishable product?

  - Who has responsibility for validity and meaning?
    lab or research community

- **Site-based dimension** - within and across geologic features
  (multidisciplinary geobiology)

Spatial – temporal vs. x3 coordinates + altitude at vent

Producer / consumer dimension

complex sets
vs.
usable parts

version
willing to release
vs.
best for reuse

DataConservancy

NSF

Curation Profiles Project

INSTITUTE of Museum and Library SERVICES

Releasable products, meaning for reuse

Use profile for NASA data levels

Raw data	Telemetry data with data embedded	**Little use** to most of science community, except radio science
Level 0 **Edited**	Corrected for errors, split into data set for instrument; tagged with time & location	**Wide use**, especially researchers familiar with instrumentation
Level 1A **Calibrated**	Corrected and expressed proportional to some physical unit	
Level 1B **Resampled**	Resampled and possibly calibrated; can't be reconstructed	**Wide use**, especially for secondary users
Levels 2-5 **Derived**		General way information transferred

Ball, A. (2010). Review of the State of the Art of the Digital Curation of Research Data.

Access & use across disciplines

Impact on research progress (cases studies in neuroscience)

- <u>Specific</u> information from <u>another field</u>
  – greatest impact / uncommon
- <u>Exploring</u> other fields
  – moderate impact / frequent

- Background and context for valid application

Rapid review and assessment - cues for "strategic reading"
(Renear & Palmer, 2009)

geobiology – image (site & conditi

Cross-disciplinary value

Complications with large-scale, quickly growing federations

(Europeana, US national cultural heritage)

small, unique & valuable flooded out by large, homogeneous

representing the whole – coverage, density, gaps

Value indicators for data

Climate / Ocean modeling
Soil Ecology
Volcanology
Stratigraphy
Sensor Engineering

- Reputation of data collector
- Spatial coverage
- Longitudinal coverage
- Site factors:
  unique conditions, rarely studied,
  permitting requirements
- Multiple sources for triangulation / context
- Documentation of workflows and provenance

Implications for NDS

- Requirements for reusable, high value data tightly tied to the paradigms and disciplinary strengths and roots.

  Particular challenge for institutional repositories

- Invest in disciplinary dimensions, value, functionality

- Metrics of success that account for impact

Acknowledgements

Data Conservancy
Data Practices Team

Tiffany Chao
Nic Weber
Karen Baker
Andrea Thomer

Melissa Cragin (now at NSF)

Site-Based Data Curation

Co-PIs: Bruce Fouke, IGB, Un. of Illinois
Sayeed Choudhury, Johns Hopkins
Ann Rodman, National Park Service
Team: JHU: Tim DiLauro; Illinois:
Andrea K Thomer, Abigail E. Esenam,
Karen S. Baker, Karen Wickett, Jacob Jett

Curation Profiles Project

PI: Scott Brandt, Purdue University

DCERC

Co-PIs: Mary Marlino (NCAR)
Suzi Allard and Carol Tenopir (UTK)
Matt Mayernik, Karon Kelly, Cheryl Thompson

Data Conservancy

PI: Sayeed Choudhury, Johns Hopkins University