INFRASTRUCTURE, DATA SHARING, AND DATA REUSE IN THE SCIENCES: THE CASE OF DATAONE

OVERVIEW

- Dissertation Research (proposal in progress)
- Mixed-Methods Approach
- Test Environment, DataONE
- Examines infrastructure factors that facilitate or interfere with data sharing and reuse
- Provides feedback for creating cyberinfrastructure for data sharing and reuse

BACKGROUND LITERATURE AND RATIONALE

Data Sharing and Rationale

Current Studies

- Scientists believe that data sharing and reuse is important
- Describes motivations and inhibitors such as journal policies
- Further examination needed

Behavioral/Technical

- Many studies conducted in the biological and biomedical domains
- Studies within the earth sciences will provide a different perspective, as scientific practices differ
- Focuses mainly on human behavior

DataONE Literature

- Overviews and summaries of DataONE infrastructure and organization, particularly member nodes, coordinating nodes, investigator tools, and community outreach program
- Theoretical discussions of DataONE as an organization including the perspective of a transdisciplinary organization, uncertainty theory, collaborative data sharing network, and complexity theory
- R&D discussion of specific aspects of the investigator toolkit or member and coordinating nodes to review and refine technical aspects.

Research Questions

What factors facilitate or inhibit data sharing and reuse in the sciences?

Within the DataONE environment, what infrastructure factors facilitate or inhibit data sharing and reuse?

Acknowledgements

We would like to acknowledge the support of DataONE and the US National Science Foundation.

The author would like to thank Dr. Jane Greenberg.

PLANNED RESEARCH METHODS

Mixed-Methods Approach

- Data Profiling Assessment
- Transaction Log Analysis
- Quasi-Experiment
- Survey
- Semi-structured Interviews

Information Gathered

- What types of data?
- What types of data are relevant?
- What types of information needed about the data?
- What types of data are being deposited?
- Which agencies depositing data?
- Which disciplines depositing data?
- What types of metadata are they providing along with their data?
- What types of searches are relevant?
- What types of information needed about the data?
- What types of data are being searched?
- What search results are relevant?
- What types of information needed about the data?
- What types of searches are relevant?
- What types of information needed about the data?
- What types of data?
- What types of information needed about the data?
- What information is important in making decisions about reuse?
- What infrastructure elements facilitate or inhibit data sharing and reuse?

PLANNED DATA ANALYSIS

- Data Profiling Assessment
- Transaction Log Analysis
- Quasi-Experimental Study
- Survey
- Semi-structured Interviews

Information Gathered

- Nuanced information regarding elements that facilitate or inhibit data sharing and reuse
- Data deposition practices of the Member Nodes

NEXT STEPS & TIMELINE

- Summer 2014
  - Data Profiling Assessment
  - Transaction Log Analysis
- Fall 2014
  - Quasi-experiment
  - Survey distribution
- Spring 2015
  - Semi-structured Interviews
  - Data Analysis
- Spring 2015 - Summer 2015
  - Data analysis
  - Writing of dissertation

Angela P. Murillo

School of Information & Library Science
University of North Carolina at Chapel Hill