Metadata Tensions: library principles vs. everyday scientific data practices
Matthew S. Mayernik, Jillian C. Wallis, Christine L. Borgman
UCLA - Department of Information Studies

Introduction: Research Data Curation

Research Data
- Research data are becoming increasingly important as scholarly products.
- The National Science Foundation now requires research proposals to include data management plans.
- Few established practices exist for managing and curating research data in large scale over long periods of time.
- Data management and curation involves technologies, policies, metadata, and effective day-to-day practices.

Metadata
- Data collections cannot be managed and curated without metadata.
- Cataloging and metadata creation are core to library services, and are based on decades to centuries of established practices and standards.
- Metadata practices for describing digital data are far less standardized.
- Institutions and infrastructures for

Problem Description: What are the metadata needs for CENS data?

What are CENS data?
- Field-based research with many different kinds of data
- Individual/lab data collection and storage practices vary widely
- CENS data resources are spread around many different computers, servers, and labs
- Many potential uses and users outside of the immediate CENS community

What is metadata?
- “Data about data” – not very helpful
- Metadata is constructed
  Not found in nature, an artificial creation.
- Metadata is constructive
  Created for a purpose, activity, or to solve a problem.
- Metadata is actionable
  Intended to be useful in some way.

Proposed Solution: Investigate day-to-day metadata and build a metadata registry

Day-to-day metadata practices
- What metadata do CENS researchers create?
  - Hand-written notes in notebooks
  - Comments in excel files
  - Headers in data files
  - Word documents
  - Comments in software
  - Web pages, including wikis and lab web sites
  - Email, both personal and mailing lists
- Why do CENS researchers create metadata?
  - Document equipment
  - Document software
  - Write up methods
  - Describe data files (column headings, units, formats)
  - Document field activities (installations, calibrations)
- How is metadata used?
  - To enable understanding and use of data
  - To share research methods and equipment
  - To create a record of work done
  - To note where anomalies occur

CENS Metadata Registration System
- We developed a data registration system to collect metadata descriptions and allow data users to:
  - find what data exists
  - Whether it might be useful to them.
  - how to get access to data
- Using a Dublin Core based schema
  - Standardizes metadata
  - Provides flexibility in fields and content
  - Is widely used in many settings
- First implementation in Spring of 2010
  - Alpha test was mixed success
  - 17% response rate, but
  - First registry and reporting of CENS data products

Metadata tensions identified
- Who is responsible for metadata?
  The responsibility for metadata creation can be ambiguous. Responsibility is divided among information professionals, working scientists, and automated hardware/software means.
- Library principles vs. metadata practices
  Libraries use a highly principled metadata approach. Everyday practices of working researchers are much more ad hoc.
- The social nature of metadata
  Metadata creation and knowledge are distributed socially in research settings. Different members of a research team have knowledge of different parts of a project.
- Where does metadata fit within the data life cycle?
  Metadata needs vary throughout the timeline of a project because research activities and data collected vary throughout a project. Approaches to collecting and organizing metadata need flexibility.