Data Management for Collaboration, Access, and Interoperability

Karl Benedict
Plato L. Smith II
University of New Mexico

CLIR Webinar – February 11, 2015
Contents

• Introduction
• Research and Data Lifecycle Models
• Data Management Considerations
• Data Interoperability and Linked Open Data
Introduction
Outline

• Data Management Planning – Foundation Principles
  • Context – Data Management Requirements
  • Relationships Between the Researcher and Data Lifecycles Models – Part 1
  • Relationships Between the Researcher and Data Lifecycles Models – Part 2
  • Data Interoperability and Linked Open Data
Context – Data Management Requirements

• Data Management Plans
• Data Sharing Requirements
• Institutional Review Board (IRB) Protocols
• Interdisciplinary Collaborative Research
• Data Intensive Research
Research and Data Lifecycle Models
Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)


- 2001 CCSDS 650.0-R-2 (Red Book)
Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)

- 2003 e-Science Curation Report (Lord & Macdonald)
Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)

Producer Perspective

- 2006 USGS (Govoni and Gunther)
Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)

Consumer Perspective

- 2006 USGS (Govoni and Gunther)
Relationship Between the Researcher and Data Lifecycles Models – Part 1 (circa 2001 to 2011)

- 2007/2015 DCC Curation Lifecycle Model
Relationship Between the Researcher and Data Lifecycles Models – Part 2 (circa 2013 to Present)

2013 JISC Research Lifecycle Diagram
Relationship Between the Researcher and Data Lifecycles Models – Part 2 (circa 2013 to Present)

- 2013/2014 Mapping Research Lifecycle to Data Lifecycle (Benedict)
Data Management Considerations
Some Definitions

• Data
• Data Curation
• Documentation (Metadata)
• Open Access
  • Consent to Share
• Embargo
  • Consent to Restrict
• License
• Data Repositories
• Long-term preservation
  • Standards
Some Recommendations

• What do you need to know?
  • *Four Kinds of Expertise
    • Domain (Subject)
    • Analytical
    • Data Management
    • Project Management
  • Professional Development and Training

• Data Assessment
  • Organization
  • Structure/Content
  • Formats
  • Documentation


http://www.dlib.indiana.edu/~jenlrile/metadatamap/
Overall Recommendations

• **Procure assistance** – consult early and often (e.g. collaborate, network)

• **Maintain documentation** from the project planning stage and throughout your work

• **Adopt a systematic model** for organizing your data: naming, file structure, formats, storage, backups

• **Adopt consistent and documented data structures**

• **Always have the entire data and research life-cycle models in mind** when you are managing your data
Data Interoperability and Linked Open Data
The Semantic Web isn't inherently complex. The Semantic Web language, at its heart, is very, very simple. It's just about the relationships between things.


... the most important thing that was new was the idea of URI -- or URL [it was UDI back then, universal document identifier]. The idea that any piece of information anywhere should have an identifier, which will not only identify it, but allow you to get hold of it. That idea was the basic clue to the universality of the Web. That was the only thing I insisted upon.

Definitions

• Interoperability
• Linked Open Data Models
• Internet Standards
  • Web Services (REST, SOAP)
• Domain Specific Standards & Protocols
  • Open Geospatial Consortium (OGC) Web Map, Web Feature and Web Converage Services (WMS, WFS, WCS)
  • DataONE, CUAHSI

An Illustration
Your Data

• How does your personal experience with data management match these goals?

• What have we learned from our experiences that can inform how we communicate with and support the researchers with which we collaborate?
Acknowledgements

• CLIR Postdoctoral Fellowship Program in Data Curation at the University of New Mexico
• NSF EPSCoR Program (Track 1 [Awards: 0447691, 0814449, 1301346] and Track 2 awards [0918635, 1329470])
• New Mexico Resource Geographic Information System
• NASA ACCESS Program
• UNM’s College of University Libraries and Learning Sciences

Data Management for Collaboration, Access and Interoperability by Karl Benedict & Plato L. Smith II is licensed under a Creative Commons Attribution 4.0 International License.