A green light to greatness.™

UNT

UNIVERSITY OF NORTH TEXAS™
Research Data Management in Academic Libraries

Meeting the Challenge

Spencer D. C. Keralis
Director for Digital Scholarship, Research Associate Professor
@hauntologist | spencer.keralis@unt.edu
#TXLA13 | #DMLib
• What we talk about when we talk about Data
• *The Problem of Data*
• Research in Data Curation Education
• Research in Data Management Policy & Practice
• Why Open Data?
• Why Libraries?
• What we talk about when we talk about Data

• *The Problem of Data*

• Research in Data Curation Education

• Research in Data Management Policy & Practice

• Why Open Data?

• Why Libraries?
Research Data Management

• NIH Data Management Plan Mandate (2003)
  – >$500K requires Data Sharing Plan
• NSF Data Management Plan Mandate (2010)
• CDC, DoE, DoD, EPA, NASA, NEH-ODH, NIST, DoAg, DoEd all have DM Plan Requirements
• White House OSTP Memo, February 22, 2013
  – digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze.
Data is: everything necessary for reproducible science.

Woods Hole Oceanographic Institute
Data is: everything
DCC Data Life Cycle Model
http://www.dcc.ac.uk/resources/curation-lifecycle-model
Technology/Infrastructure

- Prepare
- Plan
- Preserve
- Publish
Scholarly Communications

- Project to Publication vector well known in Libraries – Open Access Repositories, SC Librarians, IP, Citations, etc.
- What about the data?
- Competing interests: Must ensure IP, citation, access, embargoes, human subjects privacy, etc. etc.
- Who pays?
• What we talk about when we talk about Data
• *The Problem of Data*
• Research in Data Curation Education
• Research in Data Management Policy & Practice
• Why Open Data?
• Why Libraries?
The Problem of Data

- U.S. LIS & iSchools curricula woefully inadequate for training data professionals.
- Lowest-ranking members of research teams responsible for data curation, without training or material support.

http://www.clir.org/pubs/reports/pub154
Few researchers are aware of the data services that the library might be able to provide and seem to regard the library as a *dispensary* of goods (e.g., books, articles) rather than a locus for real-time research/professional support.

• What we talk about when we talk about Data
• The Problem of Data
• Research in Data Curation Education
• Research in Data Management Policy & Practice
• Why Open Data?
• Why Libraries?
Less than 8% of the [LIS] courses qualified as data centric. Based on course codes assigned by the schools, these courses tended to be new, and some did not yet appear in course catalogs [,] being added by institutions during the verification process. Data centric courses were rarely required for LIS degrees, certificates, or other credentials.

Given the improbability that students will encounter a data-centric course in their line of study, students must be committed to constructing a data-intensive education for themselves in order to come out of most existing U.S. LIS programs with the skills and knowledge necessary to support the needs of data-intensive research.

How can LIS education be changed and enhanced to result in well-prepared information professionals, scientists, and scholars that can take on the challenges and problems of digital curation, data management, and digital preservation, incorporating extensive training and practical experience in the context of contemporary distributed learning and web-based courses?

http://icamp.unt.edu/
• What we talk about when we talk about Data
• *The Problem of Data*
• Research in Data Curation Education
• **Research in Data Management Policy & Practice**
• Why Open Data?
• Why Libraries?
The DataRes Project, funded by a Laura Bush 21st Century Librarians grant from the IMLS, investigates how the library and information science (LIS) profession can best respond to emerging needs of research data management in universities.

DataRes is a collaboration between the University of North Texas Libraries, the UNT College of Information, and the Council on Library and Information Resources.
Project Aims

Governmental funding agency requirement

NSF  NIH
NEH  IMLS

Goals:
• Research consistent policies or programmatic implementation from research institutions
  • Emerging patterns in data management (DM) policies
• Research libraries and LIS responses
  • Curriculum and training development
• Expectations of major stakeholders in the research cycle
Digging Deeper: Text Analysis of Agency Guidance
Preliminary Findings: Institutional Policy Scan

Policy Scan Results

- 82% Institutional-Level Policy
- 18% No Institutional Level Policy
Text Analysis: Institutional Policy Scan

38 institutional policies

Example language:
“The University recognizes the importance of data sharing in the advancement of knowledge and education.”

(University of New Hampshire “UNH.VII.C.9”)

An index of known policies, with links to the project data set will be available in February 2013 at http://datamanagement.unt.edu/policies
Survey Findings

Where does your institution offer Data Management Plan Support?

- Library 48%
- Office of Research 24%
- IT Department 17%
- Individual Departments 8%
- LIS School 3%
University Interventions

• University of Minnesota – Workshops & CE Credit
• UIUC – Liaison Librarians
• Purdue & SDSU – Infrastructure & Supercomputing
• California Digital Library – Repository and DMPTool
Secondary Survey

• Intended participants: VPs of Research, Deans, High-level Administrators
  – We feel we got a very clear picture from the librarian perspective with the initial survey
  – Questions were raised through the course of our survey analysis and research that we feel only institutional administrators will be able to answer.
Key Takeaways

• Institution-level policies are last response to occur, if at all.
• Research Data Management is not a single department issue, nor is it the purview of a single discipline.
• Collaboration, domain knowledge, and infrastructure are all key to the success of any DM response.
• Prescriptive guidance from agencies will be necessary to encourage disciplines to take DM Plan requirements seriously in evaluating funding applications.
• What we talk about when we talk about Data
• *The Problem of Data*
• Research in Data Curation Education
• Research in Data Management Policy & Practice
• **Why Open Data?**
• Why Libraries?
Aaron Swartz
1986-2013
#pdftribute
The Denton Declaration
May 22, 2012 at the University of North Texas

• Open access to research data is critical for advancing science, scholarship, and society.
• Research data, when repurposed, has an accretive value.
• Publicly funded research should be publicly available for public good.
• Transparency in research is essential to sustain the public trust.
• The validation of research data by the peer community is an essential function of the responsible conduct of research.
• Managing research data is the responsibility of a broad community of stakeholders including researchers, funders, institutions, libraries, archivists, and the public.
• What we talk about when we talk about Data
• *The Problem of Data*
• Research in Data Curation Education
• Research in Data Management Policy & Practice
• Why Open Data?
• **Why Libraries?**
Library + Data = ___________

The domain of Information Sciences:

- Metadata
- Controlled Vocabularies
- Long-term preservation
- Infrastructure
- Discoverability
- Accessibility
- Reuse
- Sustainability
- Centrality/Neutrality
Scholarly Communications

- Open Access Repositories include Data associated with Publications
- SC Librarians trained in Data IP
- Develop protocols for Data Publication, Citations, etc.
- Develop models for sustainable support of Data Publication and long term preservation and access
- Buy in from OR, indirect costs
- Invest in infrastructure
The Fun Stuff

- Data Visualization
- GIS
- Apps
- STEAM in STEM
- New Media Art
- Games

Joy Division, *Unknown Pleasures* (1979)
Graphic designer Peter Saville used signal data from a pulsar for this iconic image.
LIS will need to develop stronger **partnerships** with domain researchers, informaticists, and other stakeholders in the research enterprise, to succeed at making research data an integral and enduring part of the information assets retained for science and scholarship over the long term.

SHOW ME THE DATA

http://manateestrategy.tumblr.com/
Keynote:
Kathleen Fitzpatrick
Director of Scholarly Communication, Modern Language Association
http://www.library.unt.edu/datamanagement
http://disco.unt.edu

@UNTDiSCo
@OASymposium
@DataRes