

Research Libraries and "Big Data"

Mike Furlough

Penn State University Libraries

CENDI/NFAIS Workshop

December 11, 2012 Washington, DC

Mike Furlough © 2012 | License: CC-BY

My Discussion

- Data policy and the role of libraries
- Challenges: workforce and infrastructure
- Deep collaboration, trust, and proving value

Mike Furlough © 2012 | License: CC-BY

December 11, 2012 2

Today's question....

What is the Library's
role with "big data?"

My reply....

What is the Library's role
with **any information?**

Core Library Roles

- Identify and **select** valuable resources
- **Organize** those materials
- **Describe** those materials
- **Preserve** those materials
- Provide **access** to those materials
 - **Help** others find them

"Big Research Data" Examples

- What researchers create, such as...
 - Video or other time dependent recordings of human response, physiological response, eye movement...
 - Predictive models of climate change impacts on decadal scale
 - Large-scale heterogeneous text corpora
 - The long tail of "small data"

Some jobs posted since June 2011

- Data Management Consultant
- Data Mining Consultant
- Data Research Scientist
- Data Services Librarian
- Design Data Librarian
- Digital Archivist
- Digital Collections Strategist and Architecture Librarian
- Digital Humanities Design Consultant
- Digital Records Archivist
- Manager, Data Management Services
- Research Data Librarian
- Research Data Management Coordinator
- Scientific Data Curation Specialist / Metadata Librarian
- Scientific Data Curator
- Social Science Data Consultant

Inflecting our roles

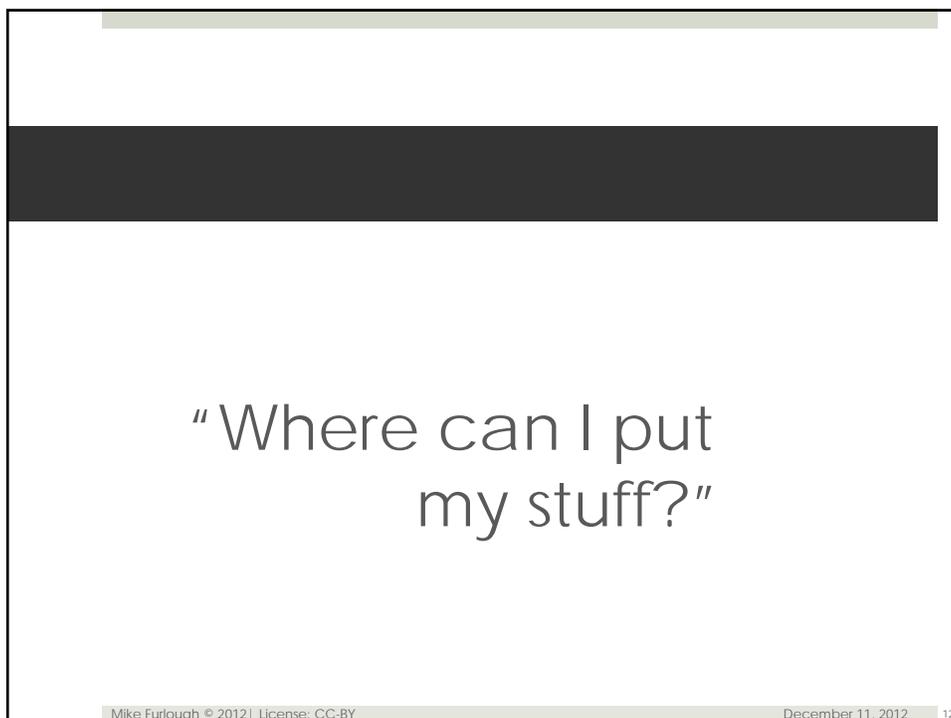
- Identification & selection, but **according to what value?**
- Organizing collections, **but of what sort?**
- Describing **according to what principles?**
- Preservation, **but in what manner?**
- Access, **but also analytics?**



What does the researcher care about?

Mike Furlough © 2012 | License: CC-BY December 11, 2012 11

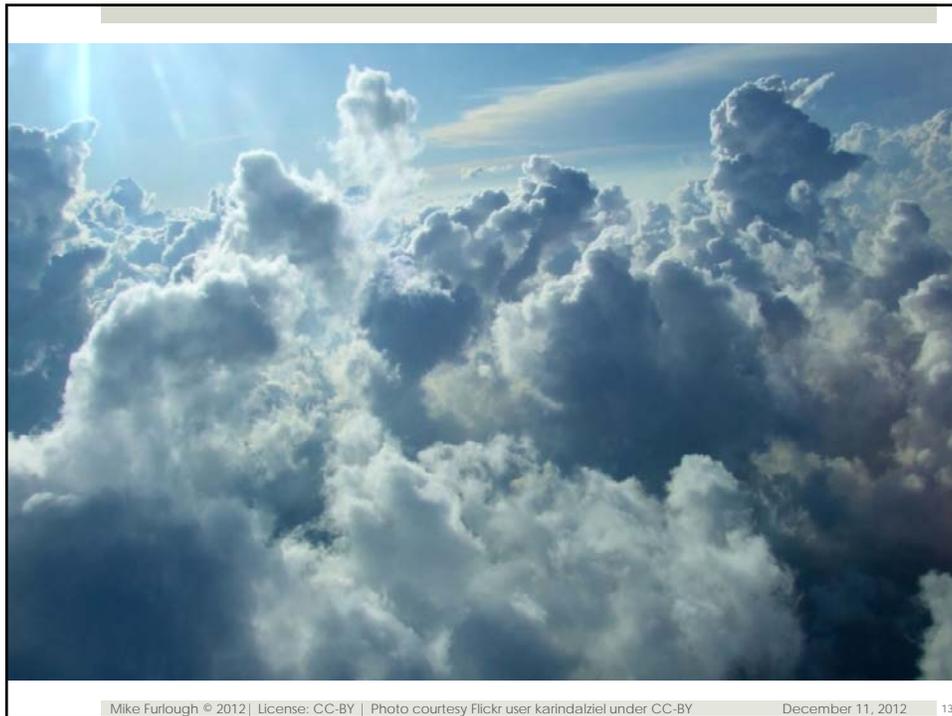
This slide features a white background with a thin black border. At the top, there is a light green horizontal bar, followed by a thick black horizontal bar. The main text is centered in a large, black, sans-serif font. At the bottom, a light green horizontal bar contains the text 'Mike Furlough © 2012 | License: CC-BY' on the left and 'December 11, 2012 11' on the right.



"Where can I put my stuff?"

Mike Furlough © 2012 | License: CC-BY December 11, 2012 12

This slide features a white background with a thin black border. At the top, there is a light green horizontal bar, followed by a thick black horizontal bar. The main text is centered in a large, black, sans-serif font. At the bottom, a light green horizontal bar contains the text 'Mike Furlough © 2012 | License: CC-BY' on the left and 'December 11, 2012 12' on the right.



The value proposition offered by libraries is longevity and durable access, not "storage."

Mike Furlough © 2012 | License: CC-BY December 11, 2012 14

Towards Deep Collaboration

- Among libraries through consortial arrangements
- Commercial-public
- With professional academic societies
- Current examples:
 - DataNet (NSF) and NDIIPP (LoC) programs
 - HathiTrust

The collage features the following logos and text:

- NDSA** (National Digital Storage Alliance)
- GeoMAPP** (Geospatial Multisite Archive and Preservation Partnership)
- DPLA** (Digital Public Library of America)
- DataCite** (Helping you to find, access, and reuse data)
- DataONE** (Digital Object Network Environment)
- Mercury** (Digital Preservation Consortium)
- DataConservancy**
- ACADEMIC PRESERVATION TRUST**
- HATHI TRUST**
- DURASPACE.ORG**
- C.I.C.** (Committee on Institutional Cooperation)
- MetaArchive cooperative**
- INDIANA UNIVERSITY**
- Data to Insight Center** (Indiana University Pervasive Technology Institute)
- DPN** (The Digital Preservation Network)

The Big Bet

"The bet is this: by curating and preserving this data we will allow for the study of 'big questions,' answers to which will benefit society. But how do you measure that? When will we know?"

Brian Schottlaender

University Librarian

University of California, San Diego

Thank you. Questions?

Mike Furlough | mfurlough@psu.edu