Research Data Management Services

Katherine McNeill
Social Sciences Librarians Boot Camp
June 1, 2012
Goals for the Workshop

• Context for research data management
• Libraries’ role in this arena
• Overview of services we provide at MIT
• Describe some specific examples of projects
• Group discussion
Context and History

- Tradition of exchanging information for scientific progress
- Importance of replication
- History of sharing of electronic data files; ICPSR celebrating its 50<sup>th</sup> anniversary
- Library data services which facilitate the access and use of secondary data
- Increased attention to research data management in all disciplines
- Includes data in any form
Reasons for Researchers to Manage their Data

• Making primary research data available can lead to further discoveries

• Funder and journal requirements:
  – NSF data management plan requirement (note: Social, Behavioral and Economic Sciences (SBE) Directorate has specific guidance)
  – NIH
  – Journal replication data requirements

• Need to curate data throughout the life cycle

• Many researchers struggle to manage their own data
Why Libraries Should Play a Role

- Evolving mission
- Greater attention to data stewardship
- Focus on curation of unique assets
- Support for faculty as information producers (in line with open access)
- Engage with others at your university
- Work further upstream in the research life cycle
- Benefits from skills enabling access to information
Data Life Cycle

History of Research Data Management Services at MIT

• Previously: formal data services with dedicated staff:
  – GIS Services
  – Social Science Data Services

• Interest among select subject librarians

• Entrepreneurial service, interdisciplinary partnership

• Reorganization: Specialized Content and Services

• Research Data Management Team
Our Services

• Web site: http://libraries.mit.edu/data-management
• Workshops
• Data storage
  – Local: DSpace@MIT (IR), Geodata Repository, and HMDC/IQSS Dataverse Network
  – Shepherd deposit in domain repositories
• Individual consultations
Individual Consultations

- Initial meeting to understand their data
- Advice on topics such as:
  - Documentation/metadata
  - Intellectual property
  - Confidentiality
  - Data conversion and file format issues
- Facilitate deposit of data in an archive or repository
- Help create data management plans for grant applications
- Referrals to other services
Outreach to Faculty on Data Management

• Identified potential faculty
• Initiated contact
• Meetings with faculty/staff
• Ongoing support

Resources from Purdue D2C2:
• Data curation profiles: http://datacurationprofiles.org
• Conducting a Data Interview, Witt and Carlson: http://docs.lib.purdue.edu/lib_research/81/
Case Study 1:
Abdul Latif Jameel Poverty Action Lab (J-PAL)

• Mission: “to reduce poverty by ensuring that policy is based on scientific evidence”
• Use Randomized Evaluations (REs) to answer questions critical to poverty alleviation
• Network of affiliated professors around the world
• Data produced: based on field collection, either surveys or administrative data; replication datasets to accompany journal articles
J-PAL Data Needs

• Location to store and make available data produced by researchers across the world
• Data should be publicly available
• Need a workflow whereby researchers can document and contribute their data in a standard way
• Solution: Repository at the Harvard-MIT Data Center, using the DDI Metadata Standard
• Ongoing data management work
Case Study 2: Shepherding Deposit at ICPSR

• MIT faculty member who ran a major survey
• Reviewed draft documentation and provided feedback
• Communicated with ICPSR regarding deposit
• Follow-up and ongoing partnership
Case Study 3: Sequencing Data

- David H. Koch Institute for Integrative Cancer Research
- Bioinformatics and Computing Core Facility
- Issues: workflows and metadata
- Opportunity to work in a new discipline
- Project just beginning
Issues to Consider

• What is your organizational culture?
• Assess the needs of researchers in your institution
• Relationship to other departments in the university
• Relationship between data specialists and subject librarians
More Issues to Consider

• To what extent do data management issues span disciplines or are discipline-specific?
• How can librarians facilitate compliance with data sharing/planning requirements?
• How can we help faculty understand the importance of data management planning?
• Determining level of service to be provided and scaling up
• Educating users to expect this service from the library
• Learning from other institutions
Further Resources

• IASSIST: [http://www.iassist.org](http://www.iassist.org)
  – Interest Group on Data Management and Curation

• Digital Curation Centre: [http://www.dcc.ac.uk](http://www.dcc.ac.uk)

• ICPSR: Deposit Data and Findings: [http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/deposit](http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/deposit)

• UK Data Archive: Manage and Share Data: [http://www.data-archive.ac.uk/sharing](http://www.data-archive.ac.uk/sharing)

Conclusion

• Build on your expertise
• Be pioneering and thoughtful
• Be proactive
• Let what faculty need be your guide
• Reach out to your colleagues in other disciplines
• Thank you! mcneillh@mit.edu
Discussion Questions

• How is your institution evolving services (or not) in this area?
• What are the needs of your faculty?
• What is your relationship with:
  – Interested librarians in other subject areas
  – Data specialists in your library?
• How does/could these services impact your relationships with faculty?
• What resources do you need to provide data management services?
• What challenges and opportunities have you been experiencing?