

RESEARCH DATA IN LIBRARIES

Rebecca Reznik-Zellen, UMASS Amherst
Donna Kafel, UMASS Medical School

NETSL Annual Spring Conference
April 8, 2011

Outline

1. What is Research Data?
2. Why Manage Research Data?
3. Approaches to Managing Research Data
 1. UMass Amherst & Institutional and Consortial Activities
 2. UMass Medical & Skills and Competencies

Disambiguation Page

- e-Science, e-Research
 - A methodology: “shorthand for the set of tools and technologies required to support collaborative, networked science.” (Hey, 2006)
- Cyberinfrastructure
 - Techno-social environment that supports data intensive, information intensive research
- Data management
 - Data management is the systematic organization and planning for data throughout the research cycle
- Data curation
 - Description and selection of data sets for long-term preservation and access

Research data is the critical component of each of these concepts

What is Research Data?

“Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results.”

University of Edinburgh Information Services

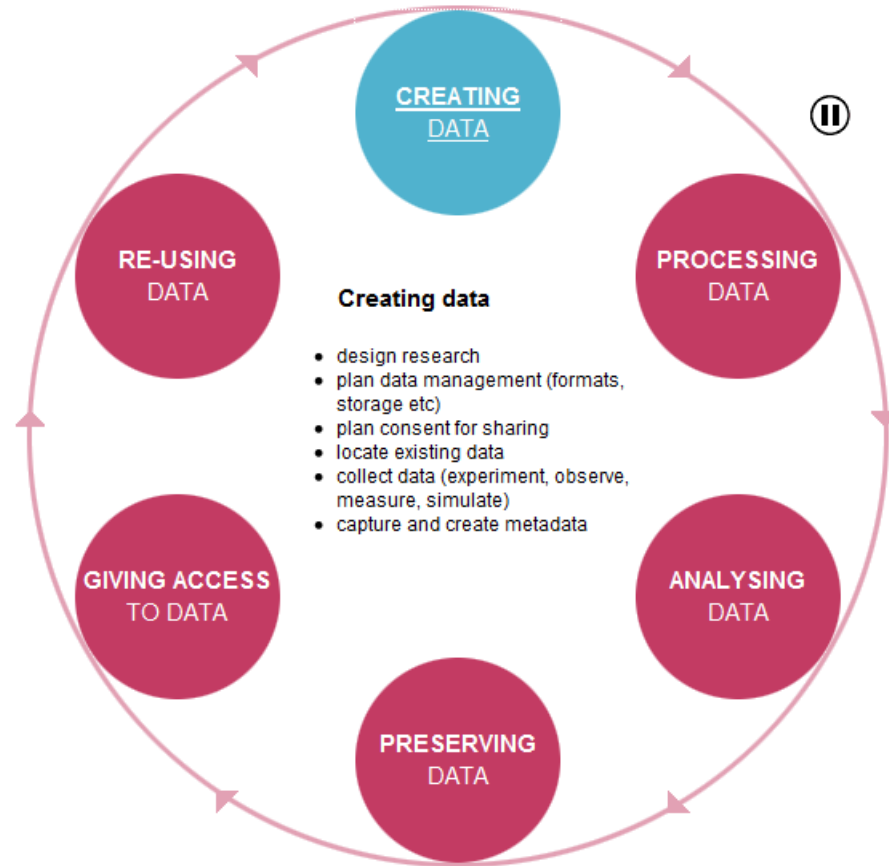
<http://bit.ly/aGXhiY>

What is Research Data?

Research publication	Research data
Information transformed into results	Information not transformed into results
Use requires basic software and instruments and their command	Use often requires special software and instruments and their command
Self-explanatory	Requires additional information and documentation if not archived
Should not include sensitive information	May include sensitive and confidential information
Use does not require permission	Use often requires permission
Ownership and copyright often clear	Ownership and copyright often unclear
Openly accessed by the scientific community for a fee or for free	Several degrees of openness (from completely open to closed)
Understood as scientific output	At the moment not understood as scientific output/merit
Ready to be used by others	Use requires processing

From Kuula (2008) Open access to and reuse of research data: the state of the art in Finland,
as excerpted in Feijen (2011) What Researchers Want

Research Data Lifecycle



UK Data Archive:

<http://www.data-archive.ac.uk/create-manage/life-cycle>

Types of Research Data...

- Observational
- Experimental
- Simulation
- Derived or compiled
- Reference or canonical



Various Formats of Research Data:

- Text
 - Numerical
 - Multimedia
 - Software
 - Models
-
- Can be specific to a discipline: crystallographic information files in chemistry
 - Can be specific to instrumentation: Olympus Confocal Microscope Data Format

Research Data May Include:

- Documents
- Lab notebooks
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, film
- Test responses
- Slides, artifacts, specimens, samples
- Models, algorithms, scripts
- Methodologies and workflows
- Data files
- Database content
- Standard operating procedures and protocols

Why Manage Research Data?

**No questions asked
\$1000 reward**

for anyone who leads to the safe return of
the stolen computer with all data intact

When: Jan 9 (Sunday) around Noon.

Where: My car (smashed car window) at Panera parking lot.

What: 13-inch White Macbook (password protected)
in dark orange computer bag.

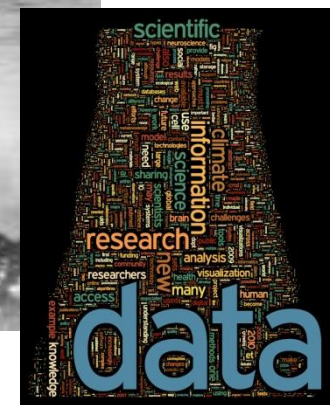
Contact: sshin2@ouhsc.edu
405-370-3099

PS. Thief, it is OK. Everybody makes mistake. Please return my computer safely
for no questions asked-\$1000 reward. If so, I would be forever grateful to you.

“Stolen laptop contains cancer cure data”

http://news.cnet.com/8301-17938_105-20028475-1.html

Why Manage Research Data?



D-Lib[®] Magazine

Why Manage Research Data?

- Protect federal investment in research and development
- Expedite the scientific process
- Use or reuse
 - the value of the data
 - the uniqueness of the data
 - the importance of the data
- Validate
- Heritage
- Obligation

Funder Requirements

- National Institute of Health (NIH)
 - Data sharing plan for grants >500k
 - Public Access Mandate for published research
- National Science Foundation (NSF)
 - Data management plan for all grant proposals
- Others: USDA, NIST, NASA...
 - See University of Minnesota Data Management Pages (<http://www.lib.umn.edu/datamanagement/funding>)



NSF Mandate

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled “Data Management Plan (DMP).” This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.

- b) “Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. “

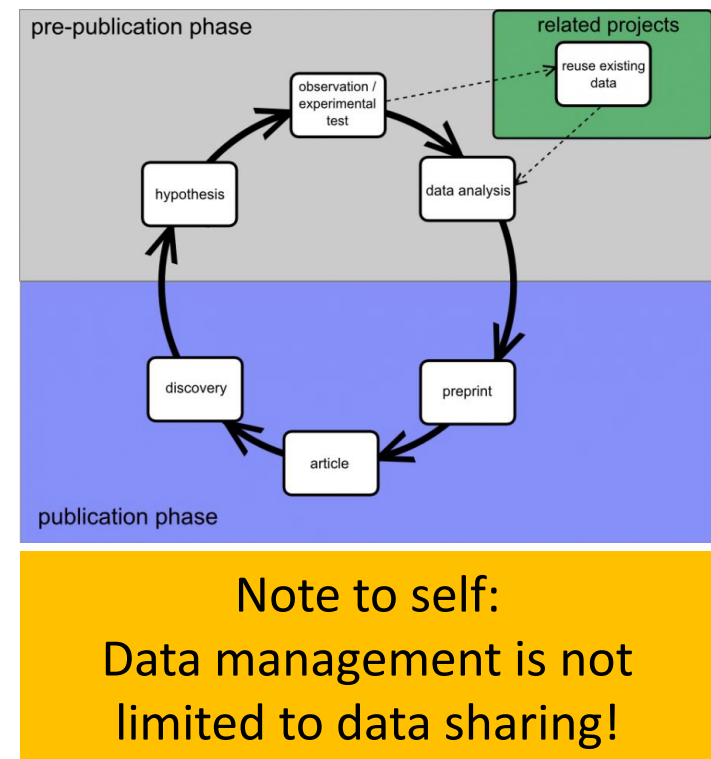
NSF Mandate

Key pieces of information to include are:

Types of data

- ☑ Metadata standards to be used
- ☑ Policies for access and sharing (including provisions for privacy/intellectual property)
- ☑ Provisions for re-use, re-distribution, and production of derivatives
- ☑ Plans for archiving and preservation of access

Data Management throughout the Research Cycle



Facilitating Compliance

Libraries are considering research data as material that falls within their scope of responsibility.

- Why

- Faculty do not want to do this
- Office of Research may not want to do this
- Neither have the expertise/infrastructure to do this

But the Libraries do: it is the libraries' natural area of expertise

- How

- By engaging faculty and crafting meaningful services to the extent that they are able

The e-Science Call to Action

e-Science presents unique opportunities to exploit and develop the capabilities of libraries/librarians.

ARL 2007, 2008

- Engaging with researchers
- Being conversant in science subjects
- Understanding nature of research methodologies and how scholarly exchange is communicated
- Understanding archival and life-cycle aspects of scientific information
- Developing standards and systems for digital content
- Data curation and preservation

UMass System Libraries' Response

- UMass 5 Science Librarians *Ad Hoc* Committee for e-Science (2008)
- Activities designed to facilitate e-Science on our campuses
 - Exploring shared electronic resources
 - Drafting Principles fundamental to e-Science
 - Creating Professional Development and Continuing Education opportunities
- A series of events (currently in third consecutive year)
 - Professional development day
 - E-Science symposium
 - Science Boot Camp

UMass System Libraries' Response

- Some principles fundamental to e-Science
 - Collaboration
 - *Curation of Primary Scientific Data*
 - *Digital Stewardship and Preservation*
 - *Metadata Standards and Creation*
 - Virtual Communities
 - Communication
 - Open Access
 - Professional Development and Investment

Principles Fundamental to the Role of the University of Massachusetts Research Libraries in e-Science

E-Science, by its nature, changes the tools that scientists use and the nature of the documentation and publication resulting from their research. Correspondingly, research libraries are changing to support e-Science in the following ways:

- Research libraries have traditionally been structured and staffed around disciplines. While e-Science embraces multidisciplinary approaches, it is worth noting that the research may have highly discipline-specific characteristics. This represents significant challenges for the research library—specializing to support particular projects and programs while addressing collaborations among traditional disciplines.
- The UMass libraries have traditionally operated autonomously, however they have collaborated on many initiatives. E-Science will provide additional opportunities for the libraries to continue their successful collaborations.
- E-Science will generate thousands of large datasets. While research libraries have broadened the scope of information formats they manage and preserve, most have not been responsible for managing scientific primary data. This is a new role for libraries, in partnership with other campus entities—stewarding the data to enable its sharing and re-use.
- Librarians must understand the disciplinary concepts, methodologies, and norms of scientific scholarly exchange. This goes well beyond the knowledge of the organization of the literature.

For the UMass Libraries to support e-Science, we must implement changes. These changes will be accomplished according to the following principles:

Collaboration

Scientific research in the information age is interdisciplinary and collaborative. To continue to support new modes of internet-enabled scientific research, the UMass Libraries are equally collaborative and will present new models of service.

Additional collaborative measures among the UMass Libraries will continue to enable knowledge, resource, and skill sharing leading to more efficient services and equity of access among campuses.

Curation of Scientific Primary Data

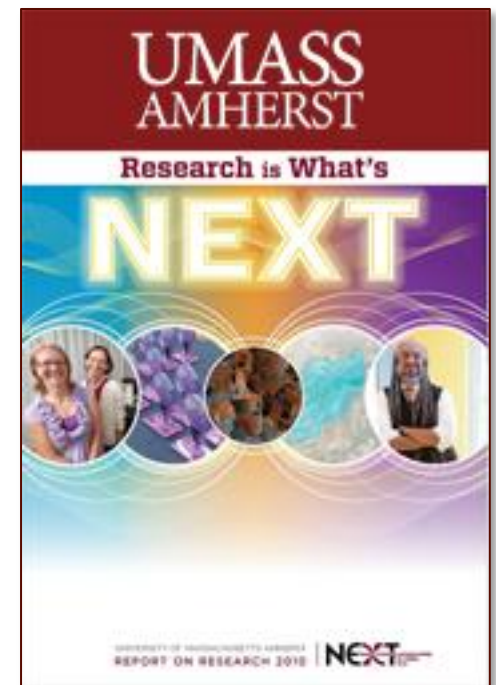
In collaboration with UMass entities, the UMass Libraries will share in the identification, management, and preservation of scientific data sets. We will provide the widest possible access to datasets, encourage repurposing, and maintain awareness of the potentially sensitive nature of some data.

Digital Stewardship and Preservation

UMass Libraries will support trusted digital repositories that meet international preservation and interoperability standards and practices, enabling the curation and preservation of a variety of research outputs from the five campus faculty and researchers.

UMass Amherst Libraries Approach

- 21,000 undergraduate students
6,000 graduate students
1,174 full-time instructional faculty
- 51 doctoral and 73 master's degree programs
- \$170 million of sponsored research in FY 2010
- 1,134 awards from 1,294 proposals submitted in FY 2010
- 498 federal awards (42% NSF)



UMass Amherst Libraries Approach

- Established a Data Working Group (DWG) to make explicit recommendations to the Libraries' regarding research data (2010)

Determine if the University Libraries should accept broad responsibility for curating research data and, if so, how that should be done, what would be expected, and who would be involved.

- Education on the issues involved with data curation
- Understanding of the University's current research environment and data outputs
- Evaluation of current Library practice for supporting active and archival data
- Exploration of partnerships for data curation both in and outside of the University
- Propose interim steps to assist the University in meeting its own policies
- Create a vision of data curation for the Libraries

Understanding the Current Environment

Faculty Interviews
and Graduate
Student Focus
Group

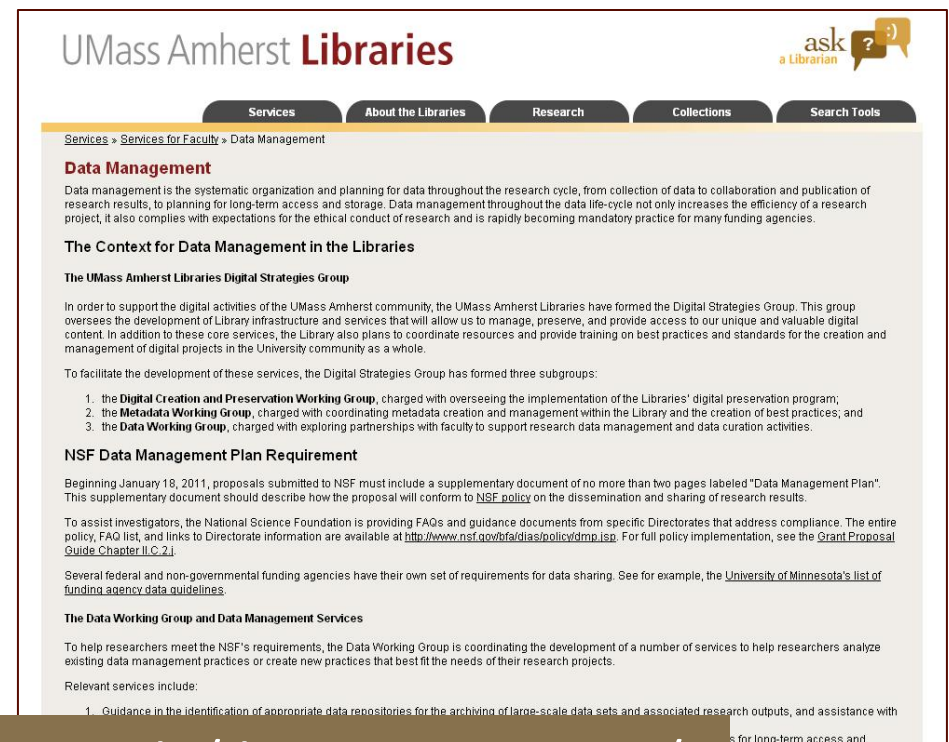
Heterogeneous
array of strategies
and infrastructure
scenarios, with
common themes
across disciplines.

Table 1: Data Management Issues by Theme

Data storage & infrastructure	<ul style="list-style-type: none">• Size• Backup/storage• Computational power• Preservation of proprietary programs/file formats• Various infrastructure ownership scenarios
Procedures & training	<ul style="list-style-type: none">• Workflow routinization• Knowledge transfer
Documentation & metadata	<ul style="list-style-type: none">• Project and discipline-specific practices• Impact of external requirement on practices
Data Reuse & sharing	<ul style="list-style-type: none">• Versioning• Making data public vs. making data useful• Collaboration for publication vs. collaborative projects• Occasional need to reproduce research
IP & data sensitivity	<ul style="list-style-type: none">• Privacy/IRB constraints• USPTO rules for confidential exposure• Emulation and Post-publication sharing

Partnerships and First Steps

- Joint letter on NSF mandate with the Office of Research
 - Outlines Services
 - Consultation on Data Management Plans
- Data Management Web Page
- Data Management Plan Template
 - Project Overview
 - Data Description
 - Data Storage
 - Access and Dissemination
 - Preservation



<http://www.library.umass.edu/data-management/>

Institutional Approaches

- Informational/Educational
 - [University of Minnesota](#)
 - [University of Nebraska](#)
- Consultative
 - [MIT](#)
 - [University of Wisconsin-Madison](#)
- Technical
 - [Purdue University](#)
 - [Rutgers](#)



Large-scale Consortial Approaches

- Linking Publications and Data
 - [Dryad](#)
- Data Management Planning
 - California Digital Libraries and Partners
- Persistent Identifiers for Data Sets
 - [DataCite](#)
- Metadata Schema for Data Sets
 - [Data Documentation Initiative](#)



UMass Medical School Approach

- 3 graduate schools: Medicine, Nursing, Biomedical Sciences
- Private and federally funded research grants > \$200 million for fiscal year 2009
- Awarded CTSA Summer 2010
- Lamar Soutter Library oversees the National Network of Libraries of Medicine New England Region (RML)
- RML outreach programs: continuing education

Evaluating local needs and competencies

- Conducted a learning needs assessment of New England science and medical librarians for planning continuing education programs and portal in 2009
- Developed an e-Science portal for New England science and medical librarians (<http://esciencelibrary.umassmed.edu>)
- Spring 2011 assessment of New England science and medical librarians' data management competencies

UMMS/WPI Data Management Project

IMLS National Leadership Planning Grant

Objectives

1. Data management curriculum for science and medical/health science students
2. Identify requirements for a data repository
3. Communications plan



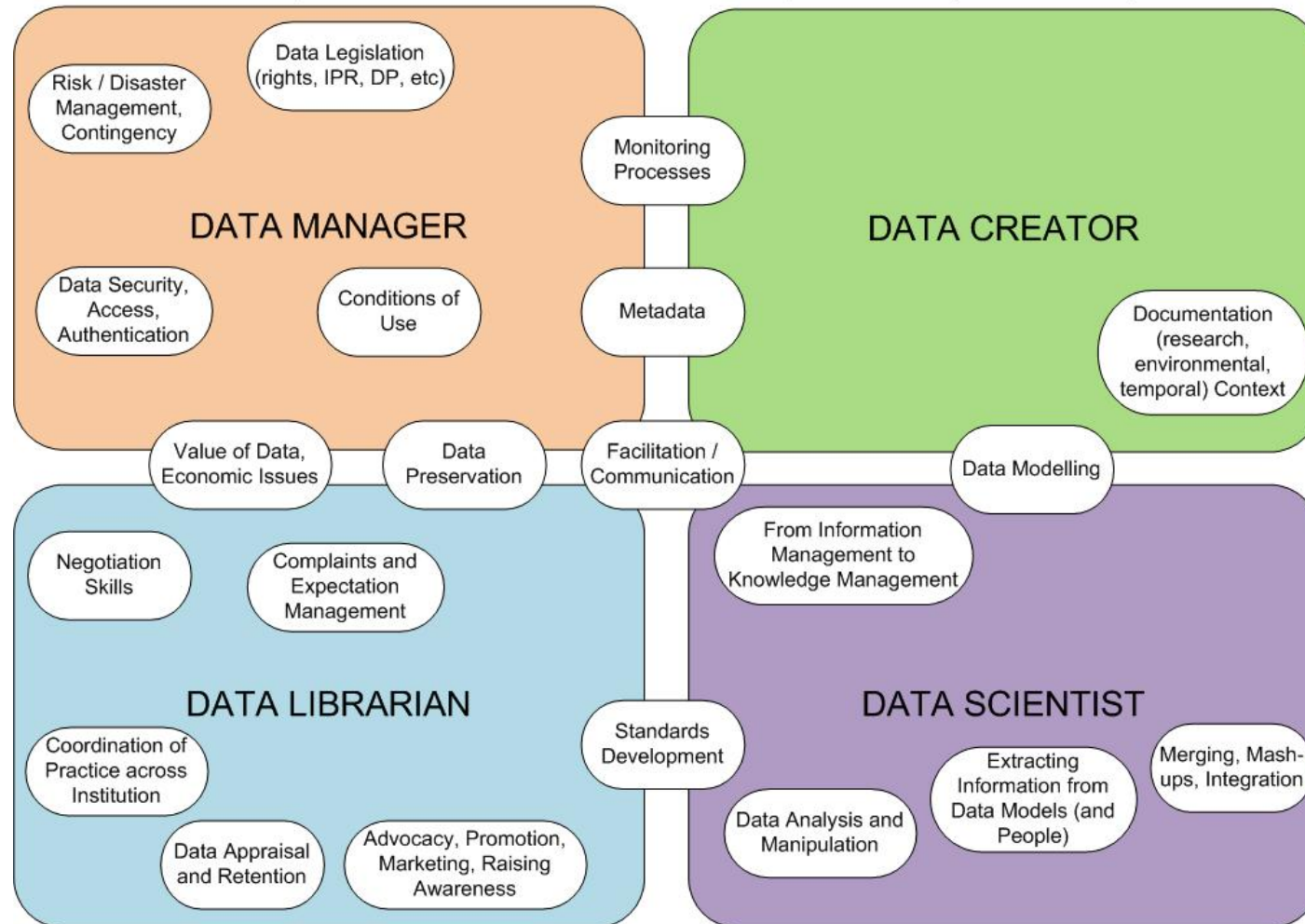
What New England libraries are doing now:

- Initiating projects to understand the scope of campus data management needs—education and environmental scanning
- Working with other campus departments to determine best practices for data management
- Developing formal policies for data management support
- Reorganizing library to have department for specialized content
- Conducting data interviews with researchers (this may be done as part of a team with subject librarians)
- Evaluating requirements for an institutional data repository
- Consulting with researchers on data management plans

Skills

CORE SKILLS FOR DATA MANAGEMENT

A follow-up from the second DCC Research Data Management Forum (November 2008)



Pryor and Donnelley. 2009. Skilling Up To Do Data. JIDC 2(4)

Data Management skills/competencies

- Understanding research methods, data lifecycle, data security
- Build, populate, and maintain digital databases
- Use a variety of programming languages (e.g. XML, SQL)
- Knowledge of metadata standards (interoperability standards, Dublin Core, MODS, OAI_PMH, etc)
- Work with metadata manipulation, crosswalk, validation, and portals
- Provide data mining, interpretation, representation, and visualization services
- Work with and develop digital lab notebook applications
- Promote digital data sharing, open access, and/or participation in IR
- Ability to work collaboratively with librarian colleagues, IT , IRB, and faculty

References

Atkins Report .[Revolutionizing Science and Engineering Through Cyberinfrastructure](#), NSF, 2003

Data Curation and Management Competencies of New England Region Health Sciences and Science and Technology Librarians
http://escholarship.umassmed.edu/escience_symposium/2011/posters/8/

e-Science Portal for New England Librarians
<http://esciencelibrary.umassmed.edu>

Feijen M. [What Researchers Want](#). Surf Foundation, 2011.

Katz L. [Stolen Laptop Contains Cancer Cure Data](#). CNet News 13 January 2011.

Kuula A and Borg S. [Open access to and reuse of research data - The state of the art in Finland. Finish Social Science Data Archive, 2008.](#)

National Science Foundation. Disseminating and Sharing of Research Results <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>

[Nature Special Issue: Big Data \(2008\)](#)

[Nature Special Issue: Data Sharing \(2009\)](#)

[The Economist Special Issue: The Data Deluge \(2010\)](#)

[Science Special Issue: Dealing With Data \(2011\)](#)

Pryor, G., and Donnelly, M. (2009). Skilling up to do data: whose role, whose responsibility, whose career? *International Journal of Data Curation*: 4(2): 158-170. <http://www.ijdc.net/index.php/ijdc/article/viewFile/126/133>

UK Data Archive
<http://www.data-archive.ac.uk/create-manage/life-cycle>

University of Edinburgh. Defining research data. <http://bit.ly/aGXhiY>