Partnering with Researchers to Share New Knowledge Through Digital Technologies

Rebecca Kennison, Director, CDRS
Partnering to Publish: Innovative Roles for Societies, Institutions, Presses, and Libraries
November 10, 2010
What Do Researchers Want from Libraries?

• Global e-access to publications, databases
• Conservation, data migration/preservation
• Metadata standards, taxonomies, ontologies
• Digitization of research materials
• Archiving, curation, dissemination of scholarly output of all types
• Technical support
  – GIS, computational data analysis, visualization, multimedia publishing
Addressing the Needs

- Information commons
- Digitization
- Digital collections, digital projects
- Virtual research environments
- Digital repositories, data repositories
What We’re Doing at CDRS

• Scholarly Communication Program
• Conferences and video services
• Journals
• Online publications
• Research repository
Scholarly Communication Program

- Information about issues (e.g., NIH)
- Coordination of copyright talks
- Administration of Columbia OA Publication Fund
- Research Without Borders speaker series
Conferences

• Caste and Contemporary India 2009
• Cities in Revolt: The Dutch-American Atlantic, ca. 1650-1815
• Columbia Undergraduate Science Symposium
• Cultural Formations
• The Digital Middle Ages: Teaching and Research 2010
• Women in Silent Film
Journals

- Awaaz: The Voice of South Asia
- Columbia Business Law Review
- Columbia Journal of Asian Law
- Columbia Journal of Race and Law
- Columbia Undergraduate Science Journal
- Consilience: The Journal of Sustainable Development
- Journal of Gender and Law
Online Book Project
Research Repository

Welcome to Academic Commons

Academic Commons is Columbia University's online repository where current faculty, students, and staff can deposit the results of their scholarly work and research. Most content in Academic Commons is freely available to the public. However, doctoral dissertations are accessible only to current members of the University community.

View recently deposited content

- Center on Japanese Economy and Business
- Economics Department
- Columbia Dissertations and Theses
- Bioinformatics and Computational Biology
- Columbia University Libraries/Information Services
- Columbia University Teachers College
- Department of Computer Science
- Department of Statistics
- Institute for Latin American Studies
- School of the Arts
- The Columbia Science and Technology Law Review
- Programs in Occupational Therapy
- English and Comparative Literature Department
- Political Science Department
- The Earth Institute at Columbia University
- Department of French and Romance Philology
- School of International and Public Affairs
- Columbia Business School
- Department of Industrial Engineering and Operations Research
- Department of Art History and Archaeology
- Department of Chemistry

Center for Digital Research and Scholarship

COLUMBIA UNIVERSITY LIBRARIES/INFORMATION SERVICES
What Is Value of Research Repository?

- Collocation
- Curation
- Interoperability
  - Consistent content models
  - Harvestable metadata for inclusion in domain- or region-oriented repositories
- Archiving and ongoing access (even when soft money dries up)
Benefits of Research Repository

- Choice of what to deposit and determination of access and reuse determined by researcher
- Research data made available alongside published outputs based on that data
- Publication (as in making public) may include negative results, incremental findings
- Value of research can be based on quality of databases, datasets, and other outputs, not on publications alone
- Data required by funders and journals to be made available or shared can be deposited in repository
- Interoperable research repositories can provide for unexpected use and novel reuse
- Impact can be tracked through robust metrics
Challenges for Research Repository

- What counts as research output varies from discipline to discipline
- Research data much more difficult to ingest, make accessible, regularize, preserve for the long-term than traditional publications; thus require much more infrastructure
- Interoperability and dynamic cross-linking of data with publications or related data not yet well-developed technologies (e.g., resource maps)
- Cooperation needed among government agencies, publishers, societies, universities, researchers
Content in Academic Commons

- Articles
- Audio/video materials
- Conference presentations
- Datasets
- Dissertations
- Monographs

- Multimedia creations
- Pre- and post-prints
- Software programs
- Technical reports
- Theses
- Working papers
Data, More Data, Even More Data

• *Journal of Neuroscience* no longer hosts “supplemental material”
• NIH’s possible strengthening of current data-sharing requirement to require public data sharing
• NSF’s data management plan requirement
  – Effective for all proposals submitted on or after January 18, 2011
NSF Data Sharing Policy

“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.”
Examples of Research Output

- Archival materials (e.g., e-mail correspondence)
- Computer executable code (e.g., simulations)
- Databases
- Datasets, spreadsheets
- Electronic portfolios
- Electronic theses, dissertations
- Lab notebooks
- Multimedia (e.g., audio, video, graphics, animations, CAD)
- Online media (e.g., blogs, wikis, Web sites)
- Photographs, images, slides
- Presentations, podcasts, pubcasts, postercasts
- Software, tutorials, documentation
- Teaching materials, learning objects
- Text files (e.g., document files, LaTeX, RTFs, PDFs)
- Visualizations
Data Management Advising

- Funding agency guidelines
- Sample data management plans
- Copyright, ethics, legal issues
- Data analysis, evaluation
- Data documentation, metadata
- Storage, backup, security
- File naming
- Data archiving standards
- Sharing, publishing data
- Citing data
- Data integration
- List of data repositories, archives
- Campus computing resources
- Campus-based data surveys
- Workshops, training, library contacts
Data Life Cycle

This model illustrates the need to manage your data at every stage of research.


This page was last updated on 07/16/09

http://libraries.mit.edu/guides/subjects/data-management/
Managing Your Data

Digital data is growing at an exponential rate: from the digital family photos on a home computer to the terabytes of data generated by researchers in the various disciplines across the university. How do we as individuals and scholars in the e-research environment keep up with our growing data management needs?

The University Libraries are working toward sustainable digital stewardship through best practices of research data management and awareness of open data issues. This site examines the life-cycle of researcher data and offers tools and solutions for creation, storage, access, analysis, and dissemination.

On this page:
Why Data Management?
Creating a data management plan
How can the library help me manage data?
beyond research data: E-scholarship

http://www.lib.umn.edu/datamanagement
Biggest Challenge: $$$

- Staffing for customization of software, education and training, curation, and data migration
- Storage: petabytes, if not exabytes
- Need for long-term institutional commitment and sustainable business models
Columbia’s Plan for NSF Data Support

• Available already
  – Long-term archive and preservation service (Academic Commons) for small datasets under 2 GB
  – Cataloging, curation for these datasets
  – Data FAQ

• For NSF grants submitted after January 18, 2011, will support larger datasets
  – Files of 2-50 GB: one-time charge of $10/GB
  – Files >50 GB: discuss needs with CDRS
Further Steps at Columbia

- Joint messaging with Sponsored Projects, Office of Research
- Continuously updated data FAQ
- Survey of data management practices on campus
- Research into possible data management services
Benefits of Deposit in Academic Commons

• Permanent URL
• Secure replicated storage
  – Multiple copies of the data, including onsite and offsite storage
• Accurate metadata
• Globally accessible repository
• Option for contextual linking of data and published research
Some Predictions

- Funders requiring mandates will increasingly pay for publication costs, data management solutions
- Research communities become even more diverse, more interdisciplinary, more geographically dispersed
- What counts for tenure and promotion will change
- Blurring of lines between traditional and new forms of communication continues
- Roles in and workflows for scholarly communication are transformed
- Search engines become increasingly better at indexing content of all types
- Semantic Web is leveraged in exciting new ways to integrate data and literature (e.g., BioLit)
Opportunities for Partnership

- Societies
- Universities
- Government agencies
- Presses
- Libraries

• How can we work together?