Dawn Paschal, Assistant Dean CSU Libraries

THE CSU DIGITAL REPOSITORY AND DATA MANAGEMENT
Outline

- Brief overview of the CSU Digital Repository
- The repository as an option for preserving, sharing data
- NSF data management plan requirements and templates
Research Goal #1

- Make your research as visible and as widely available as possible
  - Increase the likelihood you’ll be cited
  - Increase the impact of your research
  - One mechanism: the CSU Digital Repository
    http://lib.colostate.edu/repository/
A Digital Repository Is:

“A set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.”

— Clifford Lynch, Director, Coalition for Networked Information
CSU Digital Repository Services

- Permissions
- Metadata
- Uploading and posting
  - Multimedia
- Access control
- Secure storage and backup
- Promotion
- Preservation
  - Monitor format changes, migrate to succeeding format when necessary and possible
Benefits

- Increase the impact of your research through open access and dissemination
- Increase citation of your research (one study suggests open access may increase citation impact 50-250%)
- Collect, store, manage, distribute, preserve your research in one central place
- Research becomes part of the institution’s scholarly record
- Retain, control your copyright in some cases
- Registered with, crawled by Internet search engines
- Provides permanent, stable URLs for linking to, citing your research
Preservation

- The digital world’s 800-pound gorilla
- Preservation of digital assets requires:
  - Discoverability
  - Accessibility
  - Quality infrastructure
    - Preservation = a traditional library function
  - Persistence
    - Corrupted files
    - Broken links
  - Transcoding, migration of files
Preservation

In other words, storing your digital data on your hard drive or burning data on to CDs and mailing them to your mother in Florida does not constitute preservation.
Preservation

As a researcher, you need to consider:

- How long will my project last?
- How will I store my data during my project?
- How long will I store my data beyond the end of my project?
Questions Thus Far?
The Repository and Data Management Services
Your Projected Needs (2009 Survey)

Research Data Storage Needed (PBytes)

- Now: 219 PBytes
- 2 Years: 1,384 PBytes
- 5 Years: 4,914 PBytes

CSU-DR = 3 TBytes!!!
Digital Assets: Management

1. Metadata
2. User's Manuals
3. Pubs
4. Data Sets

Disciplinary Repositories, SC Centers, etc.

“Pointers”

Local Storage

Libraries-DR

Local

CSU

Small
Medium
Large

“The Cloud”
Desired Architecture

The Digital Repository: Primary & Backup Copies

High-speed Networks

LOCKSS: 3-4 Copies Distributed

CSU

UCB

CSM

UW

Preservation System
Proposed Infrastructure Model

Libraries
• Data organization & structure
• IP issues
• Metadata
• Discoverability
• Preservation

IT
• Storage capacity
• Transport capacity
• Back-up
• Sysadmin
• IT security/privacy
• Transcoding

Joint Operations

Data/Info Stewards
• The ‘front end’
• Interactions w/ researchers

System Stewards
• The ‘back end’
NSF
Data Management Plans and Templates
NSF’s Data Management Plan Requirement – Why?

- Maximize the value of data through:
  - Discoverability
  - Accessibility
  - Preservation
  - Management

- A 2008 inventory of 1,600+ federally-funded social science projects revealed at least 25% of data had been lost*

NSF’s Data Management Plan Requirement

- NSF requires proposals submitted as of Jan. 18, 2011 to include plans for data management:
  - Will contribute to merit review scores

- Other agencies have similar requirements
  - Currently many discussions about data policy at national level
Data Management Plan

Requirements Suggested by NSF

http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp

- Types of data, samples, physical collections, software, curriculum materials, & other materials to be produced;
- Standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- Policies and provisions for re-use, re-distribution, and the production of derivatives; and
- Plans for archiving data, samples, and other research products, and for preservation of access to them.

March 24, 2011

Data Mgmt.
Three templates in Word located here: http://lib.colostate.edu/repository/nsf

- Small data sets (up to 2 GB)
- Medium data sets (200 GB to 100 TB)
- Large data sets (over 100 TB)
- Data sets may be stored in the CSU DR, elsewhere on campus, in disciplinary repositories, at supercomputer centers

For each category of content suggested by NSF, a brief explanation of what you might include has been supplied

Some sample text about metadata and the CSU Digital Repository is provided that you may use in/modify for your plan as appropriate
Suggestions for Creating NSF Data Management Plans

- It’s not yet clear what NSF wants to see in data management plans; will evolve. So—
- Be specific
  - You don’t want the reviewers to make assumptions
  - Demonstrate you’re considering data management carefully
- Provide enough detail
  - Data in this context may include many things (images of glaciers, scientific papers, computer code)
  - If you are unable to provide certain types of information now, indicate they will be determined at a future date
- If certain types of data cannot be shared, mention the reason(s)
  - If you have no plans to share data, NSF wants to know why
Questions