Research Data Curation Pilots: Lessons Learned

The UC San Diego Library

David Minor, and a cast of thousands
Data Curation Pilots
The curation pilot goals
How did we begin to approach this work?

- Metadata work with researchers
- DAMS for search and discovery
- Chronopolis for digital preservation
- EZID service for object identifiers
- Training classes on Data Management Plan Tool

*Not research data!*
Metadata work with researchers

- Introductory meetings
- Metadata point person
- Ongoing discussions
- One-on-one work

Iterative, collaborative, customized, experimental...pilot!
Researchers: What is an object?

- Brain
- Slice
- Site
- Artifact

or

Etc...
Metadata lessons learned

• *Data owners are the subject experts*

• *There are many broad, basic similarities*

• *There is no such thing as one size fits all*

• *We needed a new data model that is flexible and accommodates disparate metadata from a variety of sources*
DAMS for search and discovery

Begun 10 years ago

RDF based metadata - allows for:

– multiple standards support (MADS, Premis, MIX, etc.)
– local attributes easily added
– linked data

Local and cloud based storage

DAMS Data Model needed changes to accommodate complex objects and collections
The Digital Collections website gathers two categories of content managed by the UC San Diego Library

Library Digital Collections
Digitized versions of selected collections covering topics such as art, film, music, history and anthropology

Research Cyberinfrastructure
Research data gathered by campus researchers as part of a campus-wide Research Cyberinfrastructure initiative

Currently totaling 98,345 images, audio recordings, videos, documents, and data sets, we continue to add items over time, and we invite you to search or browse these collections for research or educational purposes.

http://library.ucsd.edu/dc
Data at Redshift=1.1 (RD0025)

y-axis projection

Show metadata
Data at Redshift=1.1 (RD0025)

y-axis projection

z = 1.10, t = 5.38e+09 yr

4000x4000 y-projection 512^2 rootgrid 7-level refinement

Show metadata

Click Here
Dude!
Lessons learned

- Displaying complex objects is a challenge
- Intermixing research data and traditional library data is a (rewarding) challenge
- Important to start with an explicit and extensible object model
Data movement

Cloud Storage

Data

X₀

DAMS Storage

Staging Area
DP specific

X₁

Working Area

Project Storage

X₁

Chronopolis
Preservation Storage

X₄

DAMS Interface

X₂

X₄

NCAR

X₃, X₅

UMD

Blue box is Data Provider and RCI read/write
Red box is only RCI read/write
Simple data flow

- Basic storage
- Highly described storage
- Fully discoverable research repository
Data flow Lessons learned

- Don’t confuse our internal processes with something data owners want
- Make clear ties to their processes
- Get out of the way
Digital Preservation Across Space & Time

Spanning academic institutions and disciplines, the Chronopolis digital preservation network provides services for the long-term preservation and curation of America’s digital holdings.

Because of the ephemeral nature of digital information, it is critical to organize and preserve the digital assets that represent society’s intellectual capital—the core seeds of knowledge that are the basis of future research and education.
Digital preservation lessons learned

• *People still think preservation should be free or very cheap*

• *Many people think it’s an automatic process that just backs-up everything*
As part of an arrangement with the California Digital Library, we provide free DOIs to campus researchers.
Identifiers lessons learned

• *Not a ton of people request DOIs*

• *Those who do use a lot of them and rely on them*

• *Need to use more automatic mechanisms*
Training classes on DMP Tool

DMP Tool
Guidance and Resources for your Data Management Plan

The DMP Tool allows you to:

1. Meet funder requirements for data management plans.

Recent DMP News

DMPTool adds 100th institution!

DMPTool2 Project – September 2013

Report

Update: DMPTool 2 Features Overview

The Library
UC San Diego
**DMPTool lessons learned**

- *Initial high attendance in classes; gave way to much lower numbers*

- *Move from faculty to grad student audience*

- *Need as much customization as possible*
So ...
General lessons learned

• Our stack of services does “work”

• Every service needed tweaking (or outright changing)

• Streamline, streamline, streamline

• Tie individual services to actual steps in researcher workflow
Thank you

Paper co-authors:
- Matt Critchlow
- Arwen Hutt
- Declan Fleming
- Mary Linn Bergstrom
- Don Sutton

http://library.ucsd.edu/dc
http://rci.ucsd.edu
http://chronopolis.sdsc.edu