Managing Data at the UNT Libraries

Mark Phillips
UNT Libraries
June 4, 2014
Lay of the Land
UNT Scholarly Works Repository
The Scholarly Works Collection is home to materials from the UNT community's research, creative, and scholarly activities. It serves as UNT's Open Access Repository. This collection brings together articles, papers, artwork, music, research data, reports, presentations, and other scholarly and creative products representing the expertise in our university community.

UNT Scholarly Works aims to:

- Provide easy access to valuable scholarly and creative materials from the UNT community
- Promote discovery through effective search and navigation tools
- Secure long-term access through stewardship and preservation
- Ensure sustainability through continuing system improvements
- Showcase UNT's research and creative achievements to a worldwide audience

To learn more about UNT Scholarly Works, see our web page or contact us at unrepository@unt.edu.
UNT Data Repository
What’s Inside this Collection.

The UNT Data Repository is a central archive to provide long-term preservation and access to the research data of the UNT scholarly community. The Data Repository works in conjunction with the UNT Scholarly Works repository to ensure long-term access to the full range of research outputs from UNT.

Browse the holdings of this collection
A number of interested parties on campus
College of Information
UNT Libraries
Office of Research

Faculty?
IMLS Funded DataRes Project 2011-2013
Research Data Management: Principles, Practices and Prospects

November 2013

Published by: CLIR
A need for local services for advising faculty pre, post, and unrelated to external funding.
Team model for offering data management services
ABOUT DATA MANAGEMENT

Research Data Management

Data management is the systematic organization and planning for data throughout the research cycle which includes: 1) Collecting and processing data, 2) Analyzing data and publishing research results, and 3) Planning for long-term access and storage. Managing research data is an integral part of the research process.

Data management throughout the data life-cycle will not only help save you time, but it will help you:

- Increase the impact of your research with data citation
- Clearly document and provide evidence for your research in conjunction with published results
- Comply with sharing mandates and copyright, and ethical compliance (i.e. HIPAA)
- Preserve data for long-term access and safeguard your investment from loss
- Describe and share data with others to further new discoveries and research

The UNT Libraries are uniquely poised to provide expertise in data management planning. The UNT Libraries have taken a team approach for data management services, leveraging experience from a variety of areas in order to provide data management assistance throughout the research data life-cycle.

Data Lifecycle Consulting Team

<table>
<thead>
<tr>
<th>PLAN</th>
<th>DESCRIBE</th>
<th>PRESERVE</th>
<th>SHARE</th>
</tr>
</thead>
</table>
| Spencer Keralis  
Digital Scholarship Co-Operative | Anjum Najmi  
Digital Scholarship Co-Operative | Laura Waugh  
UNT Libraries Digital Projects | Kris Helge  
UNT Libraries Public Services |
| Consult with researchers on developing data management plans and utilizing the DMP Tool | Consult with researchers in regard to grooming research data for deposit in the UNT Data Repository | Consult with researchers and provide guidance for the deposit of data and scholarly outputs in the UNT Data Repository and UNT Scholarly Works | Consult with researchers on issues related to copyright and licensing for scholarly outputs |
Pre-Award
(I need a thing called a data management plan)
DMPTool

Training and Workshops
Post Award / Post Project
UNT Scholarly Works
UNT Data Repository
But then there is the other stuff...
Previous awards, non-grant funded, “stuff”
Data Day
Deposit-fest?
Help a faculty member through several of the hoops that they encounter while trying to deposit
Privacy Issues?
Rights Issues?
File Formats?
Data Documentation?
Descriptive Metadata?
How to I get it to you?
The goal being to get the data that they bring to the data day into the repository
Basic set of services for all deposited data
### Coda Archival Digital Repository Dataset

#### Description:
This dataset contains information extracted from the UNT Libraries' Coda Digital Repository. It contains information related to number of files, size, and ingest date of digital objects added to that system. It can be used for analysis and investigation of the growth and makeup of digital repositories.

#### Creator(s):
- Phillips, Mark Edward
- Ko, Lauren

#### Creation Date:
April 1, 2014

#### Partner(s):
UNT Libraries
- About | Browse this Partner

#### Collection(s):
UNT Data Repository
- About | Browse this Collection

#### Usage:
- Total Uses: 11
- Past 30 days: 2
- Yesterday: 0

Feedback:
If you are having problems, need to report errors, or have questions or comments for the staff, please use our Feedback Form.
Ability to download what was deposited. Easy to take and reuse.
### Coda Archival Digital Repository Dataset

<table>
<thead>
<tr>
<th>Filename</th>
<th>Size</th>
<th>Format</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>README.txt</td>
<td>1.4 KB</td>
<td>text/plain</td>
<td>Download</td>
</tr>
<tr>
<td>coda_2009-09_to_2014-03.tsv</td>
<td>43.7 MB</td>
<td>text/plain</td>
<td>Download</td>
</tr>
</tbody>
</table>

**Feedback:**

If you are having problems, need to report errors, or have questions or comments for the staff, please use our [Feedback Form](#).
Most of the tricky bits related to dataset specific metadata are pushed into a readme.txt file

(column values, sources, values, assumptions, system requirements)
Coda Archival Digital Repository Dataset:
===================================================================================

This dataset is being made available using a Creative Commons CC0 Public Domain Dedication, see http://creativecommons.org/publicdomain/zero/1.0/
The data is provided 'as is' without warranty or any representation of accuracy, timeliness or completeness.

The data format is tab-delimited text in utf8 encoding

Files:
------
coda_2009-09_to_2014-03.tsv : Dataset extracted from the Coda Archival Digital Repository operated by the UNT Libraries. Dataset ranges from September 2009 until March 31, 2014

The newspaper dataset has the following fields:
-----------------------------------------------
Coda Identifier: string used to uniquely identify a digital object in the Coda Repository
File Count: number of files that make up the digital object
Bytes: number of bytes consumed by the digital object
Ingest Date: date of ingest into the Coda Repository
External Identifier: external identifier used to link a digital object in the Coda Repository to an external system.

If you find errors or have questions do contact Mark Phillips at mark.phillips@unt.edu

Data Sources:
http://coda.library.unt.edu/ (restricted to the UNT Libraries)
The goal being to provide a basic set of services to assist faculty with their scholarship which relies on or produces data.
Questions?

mark.phillips@unt.edu

http://digital.library.unt.edu