Long tail of Research Data

Making the link from the long tail to libraries

Charles (Chuck) Humphrey
University of Alberta Libraries
2014 May
Outline

- Context represented by the long tail of research data
- The long tail applied to project-level research
- Library roles in the research & data lifecycle
  - Institutional roles
  - Project-level service roles
  - Inter-institutional network roles
Long Tail of Research Data
Long Tail of Research Data

Volume Measured in Petabytes

Numbers of Datasets
Research Data Management Infrastructure

- The combination of technology, services, and expertise organised locally or globally to support research data activities across the research lifecycle.

In Canada, we have gone from building a national institution to support research data to building research data management infrastructure. This infrastructure is being built from the bottom-up with library involvement.
The Long Tail of Research Data Applied to Project-level Research and Uses of Technology

The number of projects that consist of large volumes of data files under 1GB requires the most help with research data management infrastructure (RDMI). As the scale in data file size increases beyond 500GB, RDMI is built into the operation of the project. These projects are in need of post-project preservation of their data.
Management & stewardship

Managing research data entails the many activities dealing with the operational support of data across the stages of the research lifecycle. This involves the “what” and “how” of research data.

Data Stewardship is about the identity of those responsible for ensuring data management activities are performed to best practice levels and standards across the complete lifecycle. This addresses “who” is responsible for specific data activities.
Research lifecycle
Library Roles

- Research data management
- Data stewardship

Network

Institution

Project Research
Library and projects

- Tools, services, and expertise
  - Data management planning
  - Metadata choices: objects and workflow
  - Project file sharing
  - Data file version management
  - DOI assignment and registration
  - Data file citations
  - Predicable data and metadata flows for submission to a data repository, including file formats
Welcome.

The DMP Tool is provided by the University of Alberta Libraries to help you write data management plans.

Learn about Data Management Plans
Looking to share your research data?
Find out more about the Libraries’ research data services

Sign in
If you have an existing account.

Sign up
New to DMPonline? Sign up today.

Please note that we are currently working on CCID authentication. Please create a new DMP Tool account. You will have the option to link your existing account to your CCID when that feature becomes available.
Meeting researchers’ needs

- An easy way to share data with one, a few, or many other researchers that does not involve the use of email or Dropbox.
- A simplified approach to entering project-level metadata that can be used repeatedly with other applications.
- A one-step method of minting DOIs for data files that can be used in publications.
- A way to manage multiple versions of data files, including keeping track of changes made to the data.
- A service that helps organize data to submit for preservation processing.
UAL DATAVERSE NETWORK

A service to support project-level research on campus and to facilitate the sharing of research data.

DATAVESES
3 Dataverses

A Dataverse is a container for research data studies, customized and managed by its owner.

RECENTLY RELEASED DATAVESES
The Thomas Melville Hess Collection May 5, 2014
Geospatial Data Collection Apr 24, 2014
Research Data Collection Apr 24, 2014

STUDIES
41 Studies, 89 Files, 67 Downloads

A study is a container for a research data set. It includes cataloging information, data files and complementary files.

RECENTLY RELEASED STUDIES
Mink, Duwikibek and others by Lamont, Martha; Hess, Thom May 5, 2014
Extra tape with song; Practice conversation; Grandfather's power song; Unity speech; more practice conversation by Sam, Edward H.; Hess, Thom
Mink & his younger brother; A love song with introduction; The Bear and the Ant by Edward, Sam H.; Hess, Thom
Cripple Guts Burlington, Washington by Conrad, Emma; Hess, Thom
The All Year Around Story by Conrad, Emma; Hess, Thom

View More >
Library and the institution

- Data stewardship
  - Research data policy for the institution
  - Data deposit and dissemination agreements
  - Suite of preservation policies

- Tools, services, and expertise
  - Data curation
  - Data dissemination
  - Data preservation
Institutional policy

UAPPOL

Policies and Procedures
- A - Z Listing
- Academic Policies
- Advancement Policies
- Research Policies
- Human Resources Policies
- Finance Policies
- Administration & Operations Policies
- University Relations Policies
- Health & Safety Policies
- Information Management & Information Technology Policies
- Facilities, Properties & Equipment Policies

About UAPPOL

Contact Us

Policy Development Toolkit

Home > Policies and Procedures > Policy detail

Approval Date: 06/03/2013   Effective Date:

Research Policy

Description: 
Office of Accountability: Vice-President (Research)
Approvers: Board Learning and Discovery Committee (BLDC)
Office of Administrative Responsibility: Vice-President (Research)
Scope:

View Full Policy

PUBLISHED PROCEDURES OF THIS POLICY

- Research Administration Procedure (Roles and Responsibilities)
- Research Records Stewardship Guidance Procedure
- Research Records Stewardship Guidance Procedure Appendix B: Research Records Classification Guidelines
Institutional policy

1. Research Policy
2. Research Records Stewardship Guidance Procedure
3. Privacy Breach Procedure
4. Research Records Risk Classification Procedure
5. Research Data Management and Preservation Guidelines
### Preservation Roles and Responsibilities of UAL Positions

<table>
<thead>
<tr>
<th>Role</th>
<th>Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Curator</td>
<td>Co-ordinate Submission, SIP Quality Assurance</td>
</tr>
<tr>
<td>Data Coordinator</td>
<td>Co-ordinate Submission, SIP Quality Assurance</td>
</tr>
<tr>
<td>Digital Preservation Officer</td>
<td>Generate AIP, AIP Quality Assurance, Co-ordinate Updates</td>
</tr>
<tr>
<td>Metadata Librarian</td>
<td>Build Metadata, SIP Quality Assurance</td>
</tr>
<tr>
<td>DITL / Storage Team</td>
<td>Co-ordinate Updates, Disaster Recovery, Data Management, Manage DM</td>
</tr>
<tr>
<td>Access / Discovery Librarian</td>
<td>Co-ordinate Access, DIP Quality Assurance</td>
</tr>
<tr>
<td>Archival Storage</td>
<td>Generate DIP</td>
</tr>
</tbody>
</table>
Library and networks

- Shared data management infrastructure
  - Tools development
  - Preservation processing and storage
  - Discovery metadata exchange
The ARC Network of Canadian Libraries

Atlantic Canada

Quebec

Ontario

The Prairies

British Columbia

Preservation System

Curated Data

Data in Publications

Virtual Research Environment
Library Roles

- Research data management
- Data stewardship

Network

Institution

Project Research