Research Data Discovery

Developing a Data Catalog at the University of Maryland, Baltimore

Patricia Hinegardner, MLS
Associate Director for Resources

Na Lin, MLS
Head, Resource Development and Sharing

Meg Del Baglivo, MLS
Metadata Librarian

Hosed by ALCTS, Association for Library Collections and Technical Services
University of Maryland, Baltimore

- Seven schools including dentistry, law, medicine, nursing, pharmacy, social work and a graduate school
- Faculty: 2,785 (1,964 full-time and 821 part-time)
- Student population: 6,777
- Grant/contract awards: $667.4 million

(UMB Fast Facts: Fall 2018 statistics)
DATA IS EVERYWHERE!
UMB Data Catalog

A tool to:

• facilitate the discovery of datasets created or used by UMB researchers

• centralize and highlight what is happening at the UMB

• promote interdisciplinary collaboration by identifying common research interest

• support the process of data re-use
Background

• New York University model — developed software and a metadata schema

• Data Catalog Collaboration Project (DCCP), a cross-university initiative
DCCP

Purpose
- Exchange ideas and best practices
- Metadata standardization
- Develop the software

Collaborators
NYU, UMB, U Pitt, UNC-Chapel Hill, Duke, UVA, Wayne State, Hofstra
New members – Weill Cornell Medicine, Memorial Sloan Kettering
DCCP <cont’d>

Organization
• Institutional leads
• Working groups
  – metadata
  – outreach
  – documentation

Communication/Collaboration
• DCCP website
• Slack – collaboration tool
• WebEx - regular meetings

Hosed by ALCTS, Association for Library Collections and Technical Services
DCCP <cont’d>

Promotion at professional meetings
• Posters
• Information sessions

New model in progress
• Core group that uses the NYU data catalog software
• Institutions interested in data sharing practices
HS/HSL Data Catalog Team

- Project Lead - project management/outreach
- Metadata Librarian - creates records
- Bioinformationist - outreach
- Technology Support - web developer/programmer
- Senior Advisor - outreach/promotion
Technology

• Software available on GitHub - https://github.com/nyuhsl/data-catalog
• Operating systems - Windows / Linux
• Software – Symfony2 (Apache & MySQL), Solr
• Web-based interface
• Open Sandbox at http://dccp-test.hsls.pitt.edu/
Outreach Strategies

• Organizational
  – Research deans, program directors
  – Work with IRB and ORD to document UMB data sharing approval process
  – Promote intramural services

• Individual Researchers
  – UMB Digital Archive contacts
  – Bioinformationist
  – Research, Education and Outreach librarians for each school
Outreach Strategies <cont’d>

• Targeted promotion
  – Presentations
    • Epidemiology
    • Center for Vaccine Development & Global Health
    • Institute for Clinical & Translational Research
  – Promotional material tailored to schools
    • Handouts
    • News blog posts
Outreach Strategies <cont’d>

Address Concerns

• Easy to use template
• Provide data deposit options
• Individualized dataset request forms
• Accept research in progress
Outreach Strategies <cont’d>

Locating Data

- Databases: NIH RePORTER, PubMed Central, Scopus
- Public repositories: Dryad, Figshare, BioLINCC, OSF
- Open access publishers: PLOS ONE, Springer Nature
Anatomy of a Record

(UMB Data Catalog - online demo)
## Internal Metadata Spreadsheet (Page 2)

<table>
<thead>
<tr>
<th>Other Resources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name/Type</td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>URL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Access</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Instructions</td>
<td>Please contact Dr. Radi Masri</td>
</tr>
<tr>
<td>Access Restrictions</td>
<td>Any???</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Detailed Dataset Information</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Type</td>
<td>Empirical</td>
</tr>
<tr>
<td>Subject of Study</td>
<td>Rats (Long Evans)</td>
</tr>
</tbody>
</table>

| Equipment used to collect/create the dataset |  |
| Software used to create, collect or analyze the dataset |  |

| Dataset File Format | ???????????? |
| Dataset Size       | ???????????? |
| Data Type          | Cytokine Expression Analysis; Histological |
Workflow for Creating Internal Record

1. Start
2. Article
   - Create template
   - Enter Manage Related Entities
   - Create Record

3. Send template to researcher
4. Researcher response
   - Y: Publish
   - N: Edit

Hosed by ALCTS, Association for Library Collections and Technical Services
Lessons Learned

• Teamwork is essential: library-wide and DCCP
• Need patience and perseverance
• Metadata librarian with science background is invaluable
• IT support is critical. Developing and maintaining software
Accessing the UMB Data Catalog

• Direct URL
  https://datacatalog.hshsl.umaryland.edu

• HSHSL website – link on main page
  https://www.hshsl.umaryland.edu

• Google search: “UMB Data Catalog”

• Google Dataset Search
  https://toolbox.google.com/datasetsearch

Hosed by ALCTS, Association for Library Collections and Technical Services
Related Links

• Software available on GitHub
  https://github.com/nyuhsl/data-catalog

• NYU Data Catalog: Cataloging and Metadata Documentation - https://osf.io/9kz5a/

• DCCP: https://www.datacatalogcollaborationproject.org
Contact Information

Meg Del Baglivo
mdelbagl@hshsl.umaryland.edu

Patricia Hinegardner
phinegar@hshsl.umaryland.edu

Na Lin
nlin@hshsl.umaryland.edu
Thank You!