Let’s build one repo for data and documents with Hyrax

Martha Stuit, Nabeela Jaffer, Jose Blanco
Samvera Connect 2019, St. Louis, MO
October 24, 2019, 3 p.m. - 3:25 p.m.
Agenda

1. Overview of the repositories
2. Collaboration between the service team and IT
3. Roadmap and MVP going forward
4. Migration and other technical challenges
Deep Blue is the University of Michigan's permanent, safe, and accessible service for representing our rich intellectual community. Its primary goal is to provide access to the work that makes Michigan a leader in research, teaching, and creativity.

Finding work by U-M authors: You can browse or search the contents of Deep Blue via the options in the sidebar of every page, then link to what you find using a URL that will always work. (Available to all)

Depositing your work: We will create a collection for you, or add you as a contributor to an existing one. To get started, contact us. (Available to U-M faculty and staff)

Recent Deposits

Cyberterrorism and its Dramatic Impact on Insurance and Security Companies
Authors: Joshua (Scientific Press International Limited, 2019-11-01)
(894.1KB)

Toward a Gentle, Generalizable Framework for Scholarly Impact Conversations
Weberbach, Rebecca (2019-10-10)
(729.2KB) (67.4KB)

Beyond the "Cool" factor: the Challenge of Making Altmetrics Indispensable
Weberbach, Rebecca (2019-10-09)
(1.5KB) (64.3KB)

Research Impact & the Library
Weberbach, Rebecca (2019-10-20)
(2.1MB) (67.2KB)

Emerald Ash Borer Impact and Response on The University of Michigan Biological Station Main Campus
Caulin, Mitchell; Hazleton, Joe; Insencio, Jacob, et. al (2019)
(655.5KB)
Background

Deep Blue Documents:
- In production since 2006
- DSpace

Deep Blue Data:
- In production since 2016
- Hyrax 2.x
Deep Blue Documents: Stats

Deep Blue: Total items, files

- Files: Line graph showing an increasing trend from 20 items in 1996 to over 180,000 items by 2018.
- Items: Similar trend to files but at a lower scale.

Pie chart:
- Text: 87%
- Image: 7%
- HTML: 1%
- Audio: 1%
- Video: 1%
- Other: 3%
- Data: 1%
Deep Blue Documents: Usage Analytics
Deep Blue Data: Stats

Total works: 287
Total file sets: 6013
Total works size: 11.1 TB
Unique authors: 103
Repeat authors: 17
Unique depositors: 92
Repeat depositors: 22
Deep Blue Data: Usage Analytics
Collaboration between service and IT

- Deep Blue Unity Group (DBUG)
- Deep Blue Service Team
Deep Blue Unity Group (DBUG)

- The work of this group was to define a Minimum Viable Product (MVP) for the shared repository. This MVP was based on what features and capabilities each repo currently has.

- Content in each repository, as well as user needs and expectations, are unique.

- Deep Blue and Deep Blue Data have also developed individual approaches and strategies in their policies and services. We are doing an in-depth review of current practices to reconsider for merging.
Collaboration between service and IT

● Deep Blue Service Team
  ○ Subgroups or pairs from the strategic roadmap:
    ■ Code development
    ■ Infrastructure
    ■ Data migration
    ■ Identifiers
    ■ Policy work
    ■ Metadata
    ■ Usability
    ■ User interface and accessibility
    ■ Workflow
    ■ Usage stats
    ■ Search and discovery
    ■ Admin functionality
Any questions?
# Deep Blue Roadmapping

<table>
<thead>
<tr>
<th>Category</th>
<th>Responsible</th>
<th>Activities</th>
<th>Sep-Dec 2019</th>
<th>Jan-April 2020</th>
<th>May-Aug 2020</th>
<th>Sep-Dec 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code Development</td>
<td>Fritz &amp; Nabeela</td>
<td>Feature gap and effort analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feature development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Noah &amp; Susan</td>
<td>Monitor puma thread pool</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Document server-side ingest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Confirm Fedora’s fixity provisions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Outline token download / offloading</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Migration</td>
<td>Jose &amp; Nabeela</td>
<td>Sample data migration (50K)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Report on lessons learned</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>list of code changes required to support the data migration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifiers</td>
<td>Jake &amp; Jim</td>
<td>Altmetrics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DOI &amp; Handles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Deep Blue MVP

<table>
<thead>
<tr>
<th>STATUS</th>
<th>TYPE</th>
<th>USER STORIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Header &amp; Footer</td>
<td>MVP</td>
<td>Every page has a header and footer</td>
</tr>
<tr>
<td>Done</td>
<td>MVP</td>
<td>Header contains: Home, About, Help, Contact, and Login.</td>
</tr>
<tr>
<td>Needs Review</td>
<td>MVP</td>
<td>Footer contains: Policy and Terms of use, Questions or Contact Us by email or using the contact form, copyright statement and powered by Samvera</td>
</tr>
<tr>
<td>To Do</td>
<td>MVP</td>
<td>Library icon takes the user to the main library page</td>
</tr>
<tr>
<td>Homepage</td>
<td>MVP</td>
<td>As a Known User, I want to have an easy access to the deposit button so I can deposit my work.</td>
</tr>
<tr>
<td>Done</td>
<td>MVP</td>
<td>As a Guest User, I can only search and browse the site, and do not have access to deposit</td>
</tr>
<tr>
<td>Done</td>
<td>MVP</td>
<td>As a Guest User, I want to login with UM credentials so that I can access website as a 'Known User'</td>
</tr>
<tr>
<td>Done</td>
<td>MVP</td>
<td>As a Guest User, I want to see the featured deposits so that I can click on the deposit and see the details</td>
</tr>
<tr>
<td>Done</td>
<td>MVP</td>
<td>As a Guest User, I want to see the recent deposits so that I can click on the deposit and see the details</td>
</tr>
<tr>
<td>Needs Review</td>
<td>Expected</td>
<td>As an Admin, I want to feature a work on the work details page so that all the users can see the list of featured deposits on the homepage</td>
</tr>
<tr>
<td>Done</td>
<td>MVP</td>
<td></td>
</tr>
</tbody>
</table>
DBU Challenges:

- Data Migration
  - Users
  - Items and files
  - Relationships

- Working with Open Source Community (Samvera and Fedora)

- Policy
  - Access
  - Deposits: Mediated vs Self Deposit
  - Handles vs DOI
  - Collections configuration

- User Expectations
Any questions?
Data migration from DSpace to Hyrax

- Perl script to extract metadata from the dspace db, create yml files for each item
- Using in-house import script and the created yml files, import the records.

**Outstanding Issues**
- Include file description metadata for file, this is missing.
- Permissions of work and files not done. This is a big and complex issue. Embargo items. Items restricted on a time range basis. Items restricted by IP address.
- More rigorous metadata mapping needed.
- Handle server resolution must still to be done.
- Characterization of the files must be done as a background process.
- UTF-8 character encoding verification.
- Became obvious how necessary the short/long display in dspace is valuable.
- Mapping of collection to works not done.
Thank you!
Questions?

“Let’s build one repo for data and documents with Hyrax”

Martha Stuit
(stuitm@umich.edu)

Nabeela Jaffer
(njaffer@umich.edu)

Jose Blanco
(blancoj@umich.edu)

University of Michigan Library