

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

Present



Deep Blue Is Expanding

(Still Free, Still Citable, Still Simple to Use Anytime)

Documents

Articles, book chapters, videos, dissertations. If it represents research, teaching, and learning at U-M campuses, share or access it here.

[Enter Deep Blue](#)

Data

The underpinnings of your research, teaching, and learning are the data you collect and analyze. Share or access it here.

[Enter Deep Blue Data](#)

<https://deepblue.lib.umich.edu>



Documents

The screenshot shows the 'Home' page of the Deep Blue Documents repository. The left sidebar contains navigation links: 'Login', 'Search Deep Blue' (with a search box and 'Go' button), 'Advanced Search', 'Browse by' (with sub-links for 'Communities & Collections', 'Titles', 'Authors', 'Subjects', 'Date'), 'My Account' (with 'Login'), 'Information' (with 'About Deep Blue', 'Help', 'Contact Us'), and 'Available Now' (with 'About Deep Blue Data'). The main content area features a 'Home' header, a descriptive paragraph about the repository, and a 'Recent Deposits' section with an RSS icon. The deposits listed are:

- Cyberterrorism and its Dramatic Impact on Insurance and Security Companies** by Afshani, Joshua (Scientific Press International Limited, 2019-11-01), 894.1KB.
- Toward a Gentle, Generalizable Framework for Scholarly Impact Conversations** by Welzenbach, Rebecca (2019-10-10), 729.2KB and 57.4KB.
- Beyond the "Cool" factor: the Challenge of Making Altmetrics Indispensable** by Welzenbach, Rebecca (2019-10-09), 1.5MB and 54.3KB.
- Research Impact & the Library** by Welzenbach, Rebecca (2019-10-20), 2.1MB and 87.2KB.
- Emerald Ash Borer Impact and Response on The University of Michigan Biological Station Main Campus** by Calvin, Mitchell; Hazelton, Joe; Inosenicio, Jacob, et. al. (2019), 655.5KB.

Data

The screenshot shows the 'Deep Blue Data' repository page. The top navigation bar includes 'About', 'Help', 'Contact', a notification bell, and the email 'stuitm@umich.edu'. A search bar with the placeholder 'Enter search terms' and an 'All' dropdown is present. The main content area features a dark background with the text: 'Deep Blue Data is a repository offered by the University of Michigan Library that provides access and preservation services for digital research data that were developed or used in the support of research activities at U-M.' Two prominent blue buttons are visible: 'Browse' and 'Deposit Your Work'. Below this is a 'Featured Works' section with two tabs: 'Featured Works' and 'Recently Uploaded'. A featured work card is shown with a 3D cube icon, the title 'Retinal fundus images for glaucoma analysis: the RIGA dataset', the depositor 'shordas@umich.edu', and a list of keywords: 'Medical imaging, Optic cup, Image processing, Optic disc, Glaucoma, Optic nerve head, Image segmentation, Fundus images, Automated glaucoma screening system'.



Background

Deep Blue Documents:

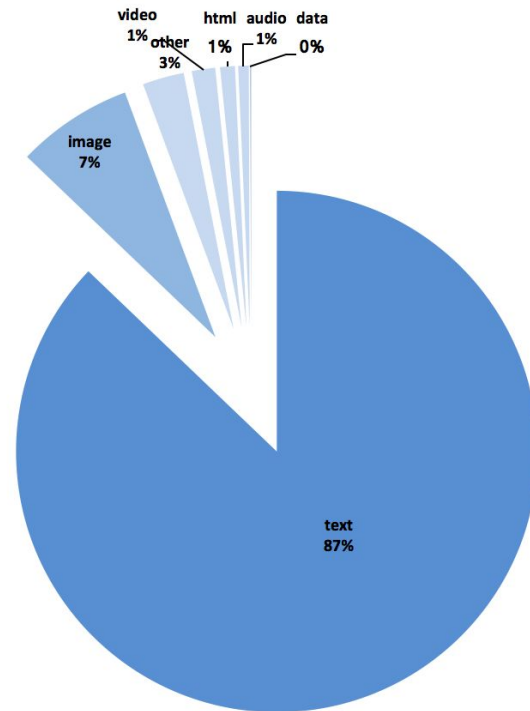
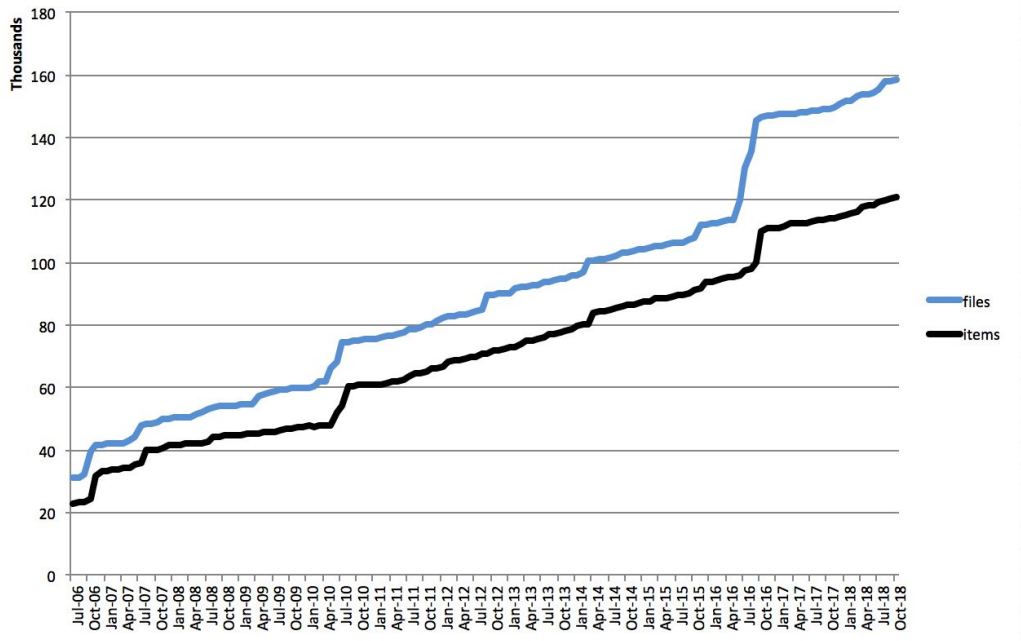
- In production since 2006
- DSpace

Deep Blue Data:

- In production since 2016
- Hyrax 2.x

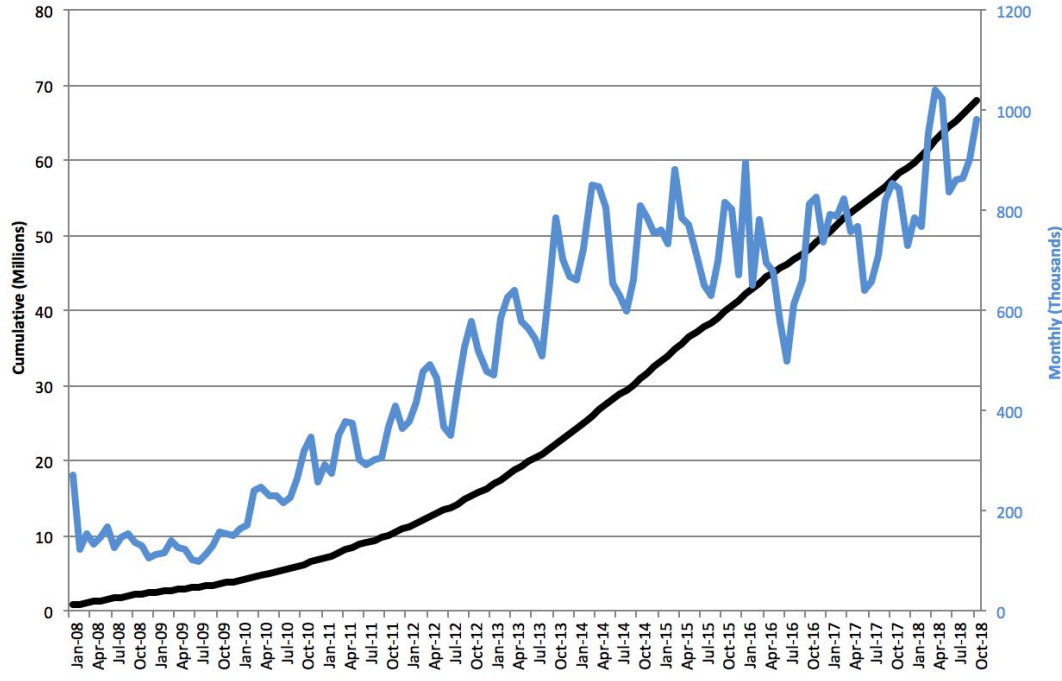
Deep Blue Documents: Stats

Deep Blue: Total items, files



Deep Blue Documents: Usage Analytics

Deep Blue: Downloads



Deep Blue Data: Stats

Total works: 287

Total file sets: 6013

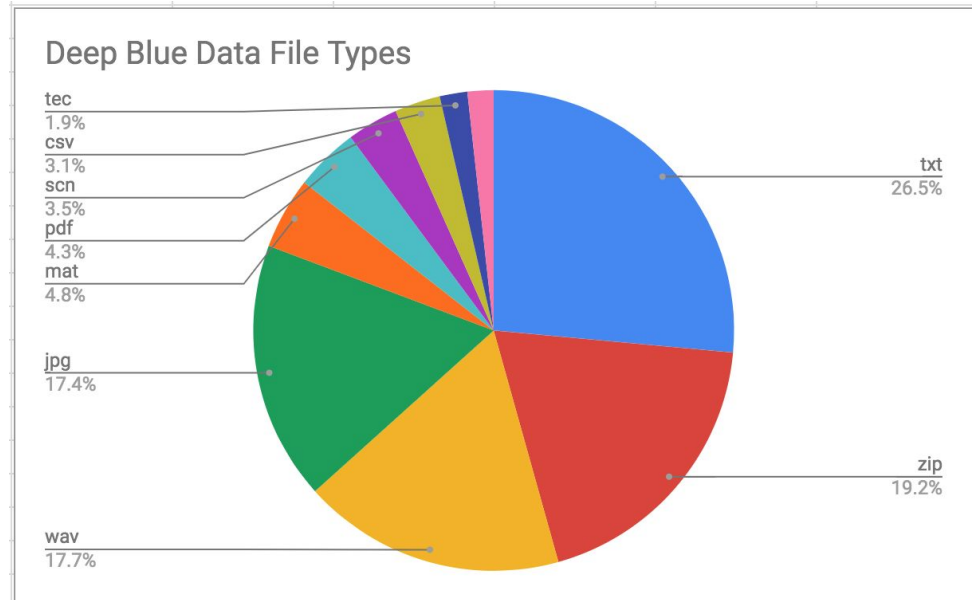
Total works size: 11.1 TB

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

A	B	C	D	E	F	G
CATEGORY	WHO IS RESPONSIBLE	ACTIVITIES	NOW / SOON / LATER MAPPING			
			Sep-Dec 2019	Jan-April 2020	May-Aug 2020	Sep-Dec 2020
CODE DEVELOPMENT						
	Fritz & Nabeela	Feature gap and effort analysis				
		Feature development				
INFRASTRUCTURE						
	Noah & Susan	Monitor puma thread pool				
		Document server-side ingest				
		Confirm Fedora's fixity provisions				
		Outline token download / offloading				
DATA MIGRATION						
	Jose & Nabeela	Sample data migration (50K)				
		Report on lessons learned				
		list of code changes required to support the data migration				
IDENTIFIERS						
	Jake & Jim	Altmetrics				
		DOI & Handles				

Deep Blue MVP

A	B	C	D	E	F
STATUS	TYPE	USER STORIES	EFFORT	DEPENDENCY	WHEN
Header & Footer					
Done	MVP	Every page has a header and footer			
Done	MVP	Header contains: Home, About, Help, Contact, and Login.			
Needs Review	MVP	Footer contains: Policy and Terms of use, Questions or Contact Us by email or using the contact form, copyright statement and powered by Samvera			
To Do	MVP	Library icon takes the user to the main library page			
Homepage					
Done	MVP	As a Known User, I want to have an easy access to the deposit button so I can deposit my work.			
Done	MVP	As a Guest User, I can only search and browse the site, and do not have access to deposit			
Done	MVP	As a Guest User, I want to login with UM crediantials so that I can access website as a 'Known User'			
Done	MVP	As a Guest User, I want to see the featured deposits so that I can click on the deposit and see the details			
Needs Review	Expected	As a Guest User, I want to see the recent deposits so that I can click on the deposit and see the details			
Done	MVP	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 deposit on the homepage			

- XS
- S
- M
- L
- XL



DBU Challenges:

**Working with
Open Source
Community
(Samvera and
Fedora)**

Data Migration

- **Users**
- **Items and files**
- **Relationships**

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

