Hyrax for Research Data
Deep Blue Data

Susan Borda - Data Workflow Specialist
Nabeela Jaffer - Program Manager

@NabeelaJaffer | @mutanthumb | #Samvera2017
What is Deep Blue Data - Hyrax based

- Expansion of Deep Blue, which preserves and provides access to the intellectual output of the UM (e.g. published articles)

- Optimized for data access and preservation

- Deep Blue Data accepts:
  - Data from all disciplines
  - Data in any format. Open, nonproprietary and widely used formats are preferable
  - Data in all sizes that are technically possible

@NabeelaJaffer | @mutanthumb | #Samvera2017
Deep Blue Data Inventory (May 2016)

- **34 deposited works**
  - Health Sciences: 16
  - Humanities: 7
  - Science: 5 (2 migrated from Deep Blue)
  - Social Sciences: 2
  - Engineering: 2 (1 migrated from Deep Blue)

- **Currently reviewing deposits**
  - Developing rubric for assessing quality
  - Will need help from subject librarians!

- **20+ “in progress” deposits**

@NabeelaJaffer | @mutanthumb | #Samvera2017
Stats 2017

- Total works: 204
- Total files_sets: 7321
- Unique authors: 30
- Repeat authors: 10
- Unique depositors: 65
- Repeat depositors: 16

@NabeelaJaffer | @mutanthumb | #Samvera2017
Stats 2017 - continued

Total Works Size Deep Blue Data (TB)

- 16.8 TB on 12/1
- 4.37 TB on 11/1
- 2.94 TB on 10/30

@NabeelaJaffer | @mutanthumb | #Samvera2017
LIT and RDS

Library IT has traditionally used waterfall approach

One of the first projects to use Agile methodology

- Improved communication resulted in less ‘do it again’
- Regular IT and RDS team meetings helped in deciding as a group
- Working demo was shared frequently to get feedback from RDS
- Flexibility in feature list while keeping the promise of meeting deadline.

@NabeelaJaffer | @pseudoAMYloid | #MichiganIT
Additional Features

Draft mode

Additional metadata

Redesigned Ingest process

Globus for big data downloads -
https://training.deepblue.lib.umich.edu/data/concern/generic_works/cv43nw83b?locale=en

@NabeelaJaffer | @mutanthumb | #Samvera2017
<table>
<thead>
<tr>
<th>File</th>
<th>Filename</th>
<th>Date Uploaded</th>
<th>File Size</th>
<th>Access</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CRMA 2017 Annual grain.csv</td>
<td>2017-11-06</td>
<td>42.7 KB</td>
<td>Open Access</td>
<td>View Details</td>
</tr>
<tr>
<td></td>
<td>CRMA 2017 Diversity x Nitrogen.csv</td>
<td>2017-11-06</td>
<td>38.8 KB</td>
<td>Open Access</td>
<td>View Details</td>
</tr>
<tr>
<td></td>
<td>CRMA 2017 Main_metadata.csv</td>
<td>2017-11-06</td>
<td>2.83 KB</td>
<td>Open Access</td>
<td>View Details</td>
</tr>
<tr>
<td></td>
<td>CRMA 2017 Main.csv</td>
<td>2017-11-06</td>
<td>70.7 KB</td>
<td>Open Access</td>
<td>View Details</td>
</tr>
<tr>
<td></td>
<td>CRMA 2017 Annual grain_metadata.csv</td>
<td>2017-11-06</td>
<td>2.98 MB</td>
<td>Open Access</td>
<td>View Details</td>
</tr>
</tbody>
</table>

Download All Files
Download All for Globus

@NabeelaJaffer | @mutanthumb | #Samvera2017
Service Side Opportunities & Challenges

Large data set challenge

Admin side

Workflow - unmediated deposit to mediation

Shared space on the server

@NabeelaJaffer | @mutanthumb | #Samvera2017
Questions + discussion

What kind of data are you working with?

- Do you want to / can you share it?
- Do you have a preservation strategy?

Image by iconsphere via Noun Project

@NabeelaJaffer | @mutanthumb | #Samvera2017