Adapting Omeka for Special Collections and Archival Content

Virginia A Dressler
Adapting Omeka for Special Collections and Archival Content

Virginia Dressler, MA, MLIS
Digital Projects Librarian
Kent State University
Background of project

• Migrated to Omeka in November 2015
  – Provided a solution that catered to previous custom interface and potential for future growth

• NHPRC grant: Sept 2016- awarded 2 year grant to digitize a large portion of the May 4 Collection (~72,000 pages of content)
Recent enhancements @ KSU

• Admin updates
  – Move to batch ingesting
  – Shift in ingest process
  – OCR processing change
  – “Auto-linker” finding aid tool

• Implemented Mirador image viewer
  – Better reflects archival arrangement
Move to batch ingest

• For the purpose of the grant, ingesting each subcollection within a larger archival collection made sense for large scale digitization

• Excel spreadsheets provide framework for metadata templates from archival collection metadata
### Sample Excel file

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dublin Core:Illustrator</strong></td>
<td><strong>Dublin Core:Publisher</strong></td>
<td><strong>Dublin Core:Date</strong></td>
<td><strong>Dublin Core:Date Copyrighted</strong></td>
<td><strong>Dublin Core:Description</strong></td>
<td><strong>Dublin Core:Local Identifier</strong></td>
<td><strong>Dublin Core:Subject</strong></td>
<td><strong>Dublin Core:Date Issued</strong></td>
<td><strong>Dublin Core:Format</strong></td>
<td><strong>Dublin Core:Institution</strong></td>
<td><strong>Dublin Core:Repository</strong></td>
<td><strong>Dublin Core:Finding aid title</strong></td>
<td><strong>Dublin Core:Finding aid URL</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Undated poem by Bill Ayres titled &quot;Guilt Trip.&quot;</td>
<td>KSU Shootings</td>
<td></td>
<td></td>
<td>archive (mixed materials)</td>
<td>Kent State University Special Collections and Archives</td>
<td>Finding Aid for the May 4 Poetry collection</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ingest individual image files instead of PDFs

• Previously, we ingested raw image files to offline storage while uploading a single PDF to reflect multi-page text items

• Offline storage is now synced to Omeka ingest
  – PDF is automatically generated upon publication (PDF/A 1-b)
  – Ingesting the image files allowed for switch in OCR processing
Tesseract OCR

• Open source software
• More accurate than Adobe Acrobat after comparison tests
• Streamline production workflow

“Auto-linker”

• Detects finding aid URL in metadata record and updates corresponding Drupal based finding aid with new Omeka link
Mirador image viewer
Future work...

• Exploring crowdsourced transcription plug-in for handwritten content
• Digital exhibits
• Expand on existing outreach and promotion of digital collections

**And all code currently available in GitHub**
Thank you!

Virginia Dressler
Digital Projects Librarian
vdressle@kent.edu