

**Kent State University**

---

**From the Selected Works of Virginia A Dressler**

---

May 13, 2017

# Adapting Omeka for Special Collections and Archival Content

Virginia A Dressler

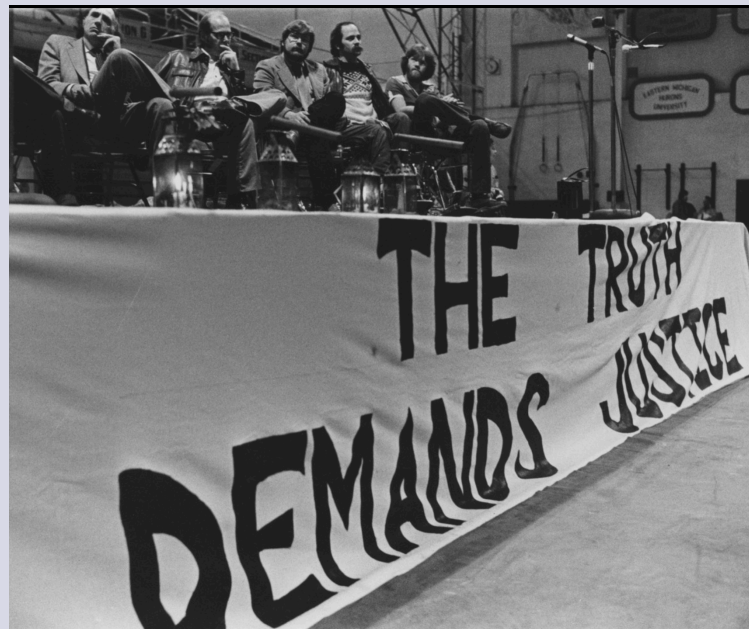


This work is licensed under a [Creative Commons CC BY-NC-SA International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).



Available at: [https://works.bepress.com/virginia\\_dressler/15/](https://works.bepress.com/virginia_dressler/15/)

# Adapting Omeka for Special Collections and Archival Content



Virginia Dressler, MA, MLIS  
Digital Projects Librarian  
Kent State University

INVESTIGATION OF STUDENTS FOR A  
DEMOCRATIC SOCIETY  
Part 2  
(Kent State University)

HEARINGS  
BEFORE THE  
COMMITTEE ON INTERNAL SECURITY  
HOUSE OF REPRESENTATIVES  
NINETY-FIRST CONGRESS  
FIRST SESSION  
JUNE 24 AND 25, 1969  
(INCLUDING INDEX)

Printed for the use of the  
Committee on Internal Security



U.S. GOVERNMENT PRINTING OFFICE  
WASHINGTON : 1969

30-133

For sale by the Superintendent of Documents, U.S. Government Printing Office  
Washington, D.C. 20402 - Price 70 cents

# 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

Dublin Core:Illustrator	Dublin Core:Publisher	Dublin Core:Date	Dublin Core:Date Copyrighted	Dublin Core:Description	Dublin Core:Local Identifier	Dublin Core:Subject	Dublin Core:Extent	Dublin Core:Format	Dublin Core:Institution	Dublin Core:Repository	Dublin Core:Finding aid title	Dublin Core:Findin
				Undated poem by Bill Arthrell titled "Guilt Trip."		KSU Shootings		archive (mixed materials)	Kent State University	Special Collections and Archives		<a href="http://www.library.u/special-collectic/archives/may-4-p/lection">http://www.library.u/special-collectic/archives/may-4-p/lection</a>
				Undated poem by Bill Arthrell titled "Springtime," dated July 10, 2006.		KSU Shootings		archive (mixed materials)	Kent State University	Special Collections and Archives	Finding Aid for the May 4 Poetry collection	<a href="http://www.library.u/special-collectic/archives/may-4-p/lection">http://www.library.u/special-collectic/archives/may-4-p/lection</a>
				Undated poem by Bill Arthrell titled "The Other Side of Blanket Hill," dated May 9, 2000. Includes a short explanation of the poem by Barnett dated May 4, 2006.		KSU Shootings		archive (mixed materials)	Kent State University	Special Collections and Archives	Finding Aid for the May 4 Poetry collection	<a href="http://www.library.u/special-collectic/archives/may-4-p/lection">http://www.library.u/special-collectic/archives/may-4-p/lection</a>
				Undated poem by Bill Arthrell titled "The Other Side of Blanket Hill," dated May 9, 2000. Includes a short explanation of the poem by Barnett dated May 4, 2006.		KSU Shootings	2 pages	archive (mixed materials)	Kent State University	Special Collections and Archives	Finding Aid for the May 4 Poetry collection	<a href="http://www.library.u/special-collectic/archives/may-4-p/lection">http://www.library.u/special-collectic/archives/may-4-p/lection</a>
				Undated poem by Bill Arthrell titled "The Other Side of Blanket Hill," dated May 9, 2000. Includes a short explanation of the poem by Barnett dated May 4, 2006.		KSU Shootings	10 pages	archive (mixed materials)	Kent State University	Special Collections and Archives	Finding Aid for the May 4 Poetry collection	<a href="http://www.library.u/special-collectic/archives/may-4-p/lection">http://www.library.u/special-collectic/archives/may-4-p/lection</a>

# 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

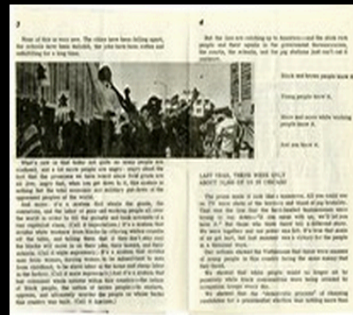
Q SDS: Publication: "Bring the War Home! Chicago - Oct. 11" (7 p...



scamay4b107f19\_...



scamay4b107f19\_...



scamay4b107f19\_...



scamay4b107f19\_...



scamay4b107f19\_...



scamay4b107f19\_...



scamay4b107f19\_...

# 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](mailto:vdressle@kent.edu)