

5-31-2024

Active and Retroactive Digital Newspaper Preservation

Andrew McDonnell
University of Kentucky, mcdonnell@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/libraries_present



Part of the [Archival Science Commons](#)

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Repository Citation

McDonnell, Andrew, "Active and Retroactive Digital Newspaper Preservation" (2024). *Library Presentations*. 272.

https://uknowledge.uky.edu/libraries_present/272

This Presentation is brought to you for free and open access by the University of Kentucky Libraries at UKnowledge. It has been accepted for inclusion in Library Presentations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu, rs_kbnotifs-acl@uky.edu.

Active and Retroactive Digital Newspaper Preservation

Andrew McDonnell
Digital Archivist
University of Kentucky
Special Collections Research Center
mcdonnell@uky.edu





• Our Commitment to the Ephemeral

- Temporal value
- Digital is the new high-acid paper stock
- Fragile Bits | Bits are Fragile
- Two University of Kentucky Case Studies

The University of _____ Kentucky's Newspaper preservation

- Microfilm, starting ~1949
- Digitization, starting ~1998
- Born-digital, starting ~2018

For a tremendous history, read Kopana Terry and Eric Weig:
https://uknowledge.uky.edu/libraries_present/137/

Case Study:

La Voz de Kentucky



Bilingual, Bi-weekly
Newspaper:
2003-2019

Optical Media

La Voz in its raw state:
362 discs, combination CD-Rs
and DVD-Rs

Sometimes labeled.
Sometimes not.



Our friend the Nimbie

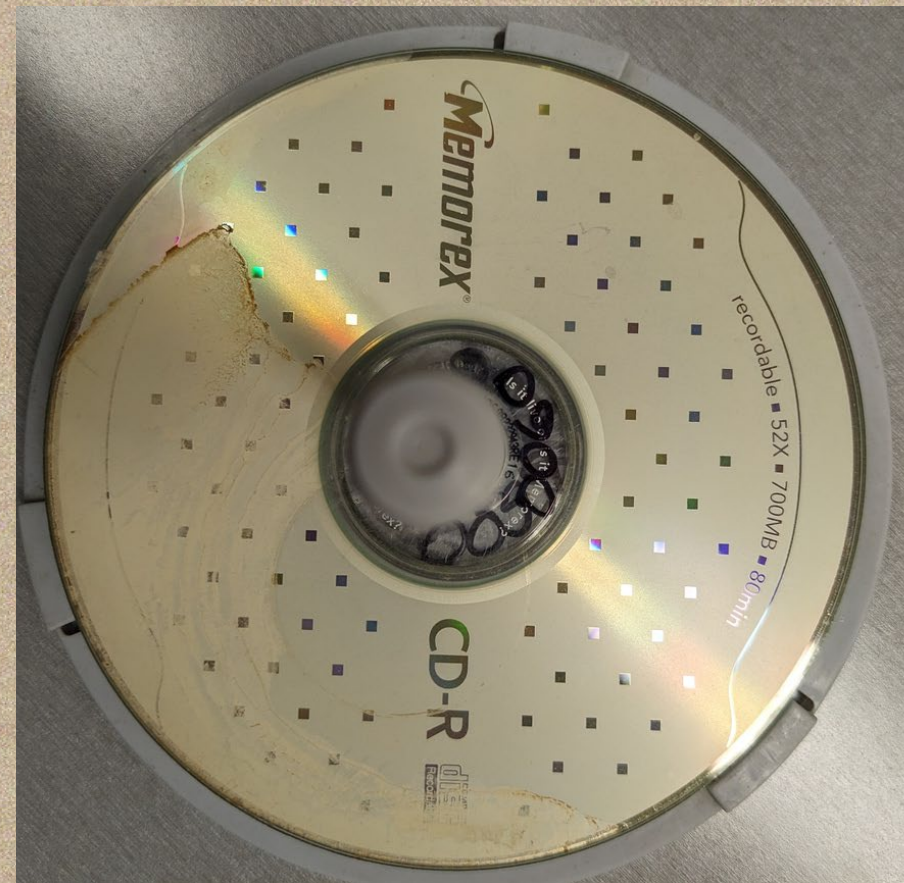


100 Discs at a Time!

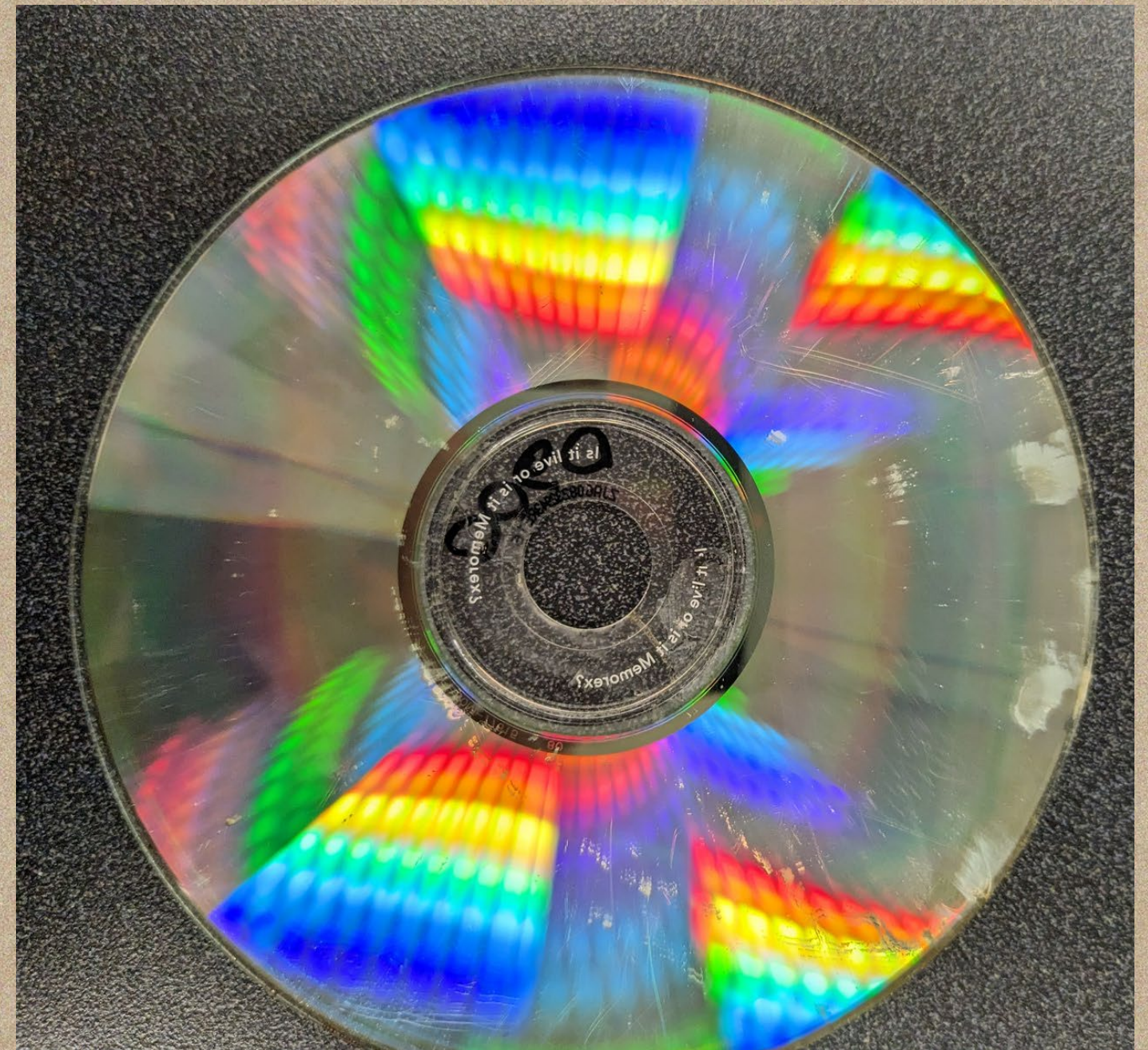
- Format agnostic software
- Skipped disc imaging, conducted straight file migration
- Rip first, ask questions later (in this case)



Our enemy: _____ Scratches and Goop



Too shiny to fully convey, but physical condition played a role in the process.



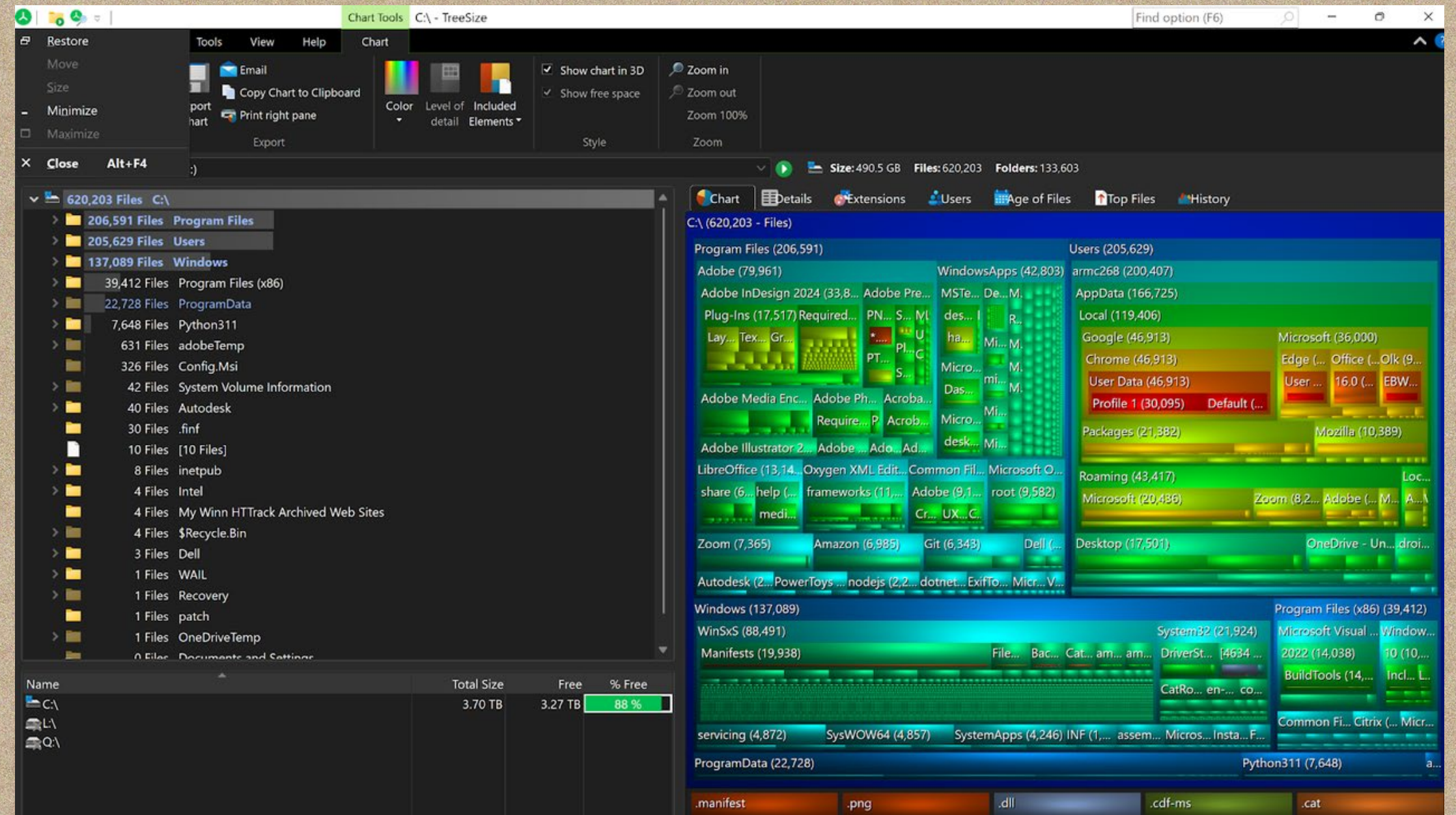
metadata preparation

1	Disc #	Transcription from case and/or disc	Contains Newspaper?	File Types	Size	Files	Folders	Notes
199	0198	Musica Mac 3	<input type="checkbox"/>	Audio/Music	660.3 MB	145	65	
200	0199	Musica 1	<input type="checkbox"/>	Audio/Music	574.6 MB	126	55	Music
201	0200	Suplemento 10-10-10	<input checked="" type="checkbox"/>	Multiple types	233.4 MB	97	3	La Voz: Vol 6 No 8: Apr 20-May 3, 2006_32 P
202	0201	Fotos Compu Andres Respaldo 2003-2005; September 26, 2007	<input type="checkbox"/>	Photograph	2.7 GB	1890	46	Photos 3 directories: 2003, 2004. 2005
203	0202	Recursos La Voz	<input type="checkbox"/>	Corrupt Disc	0.0 GB	0	0	Disc unopenable
204	0203	Blank	<input type="checkbox"/>	Photograph	101.2 MB	38	0	June- July 2007-08
205	0204	Blank	<input type="checkbox"/>	Blank disc	0 Bytes	0	0	empty
206	0205	Blank	<input checked="" type="checkbox"/>	PDF	30.5 MB	22	0	La Voz: Vol 8 No 16: Aug 7-20 2008_28 Page
207	0206	La Voz	<input checked="" type="checkbox"/>	Multiple types	12.4 MB	3	0	Logo
208	0207	Jet Set 2 Airport Lounge Riffage	<input type="checkbox"/>	Multiple types	586.9 MB	258	4	mostly mp3 files/ graphics/ mac install files/
209	0208	Adobe folders	<input checked="" type="checkbox"/>	Multiple types	127.5 MB	2192	156	La Voz: Vol 5 No 7: Mar 31-Apr 13, 2005_36
210	0209	Elmer, Espanol final	<input type="checkbox"/>	Audio/Music	88 Bytes	2	0	cda files
211	0210	Fonts P.C.	<input type="checkbox"/>	Multiple types	192.8 MB	2139	1	Fonts
212	0211	Blank	<input checked="" type="checkbox"/>	Multiple types	187.1 MB	96	3	La Voz: Vol 6 No 20: Seo 21-Oct 4, 2006
213	0212	Adobe Periodico	<input checked="" type="checkbox"/>	Multiple types	64.6 MB	5	0	Dr. Carrillo photo, Photoshop images

Post rip analysis: initial manual evaluation and tracking.

Appraisal while you work.

TreeSize Analysis



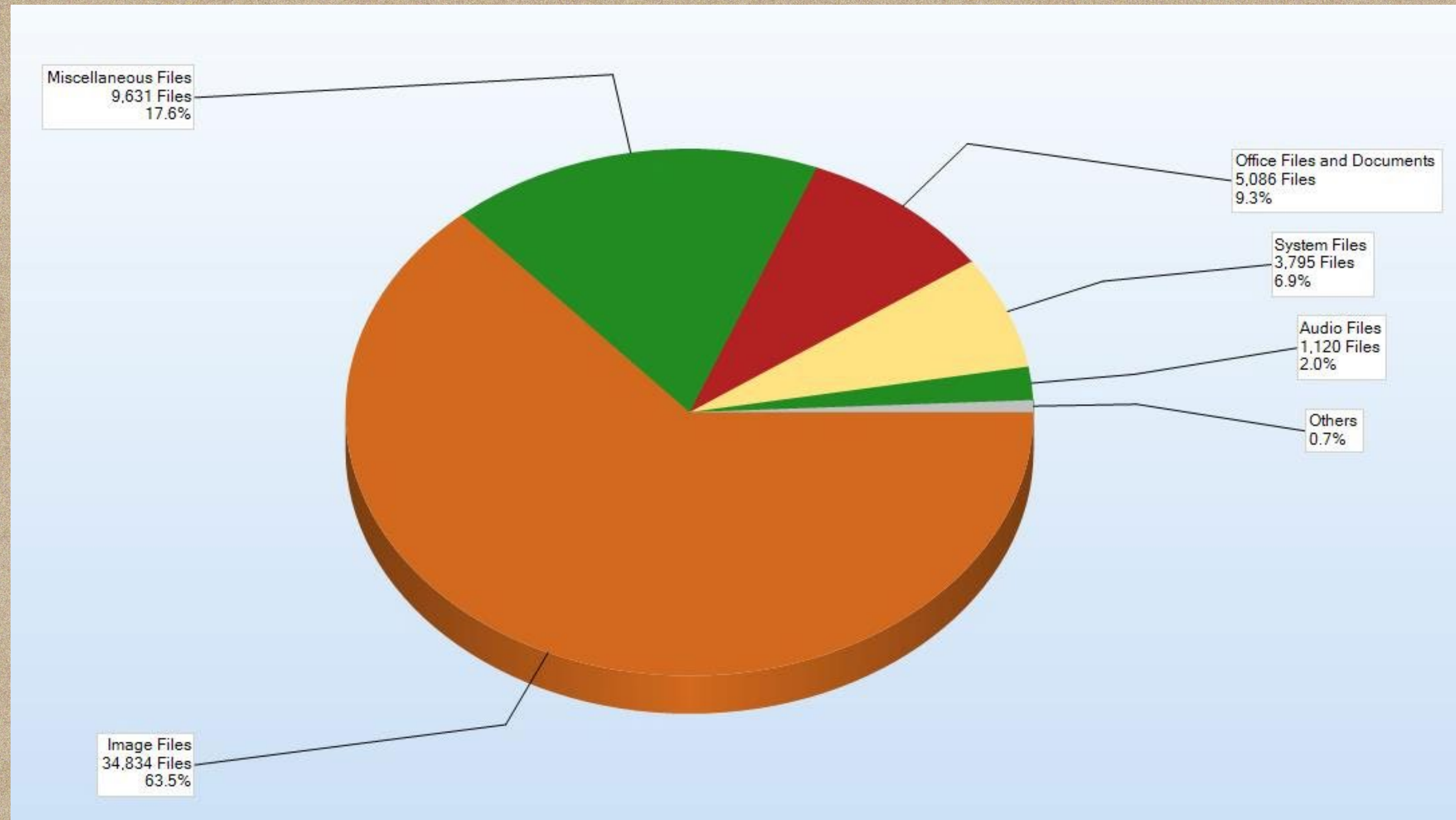
- TreeSize overall file characterization and file manifest
- File extension analysis revealed more than newspaper files in the collection
- Multiple ways of visualizing, exploring a collection

Appraisal while you work.

Treesize Analysis

Manual Intervention revealed

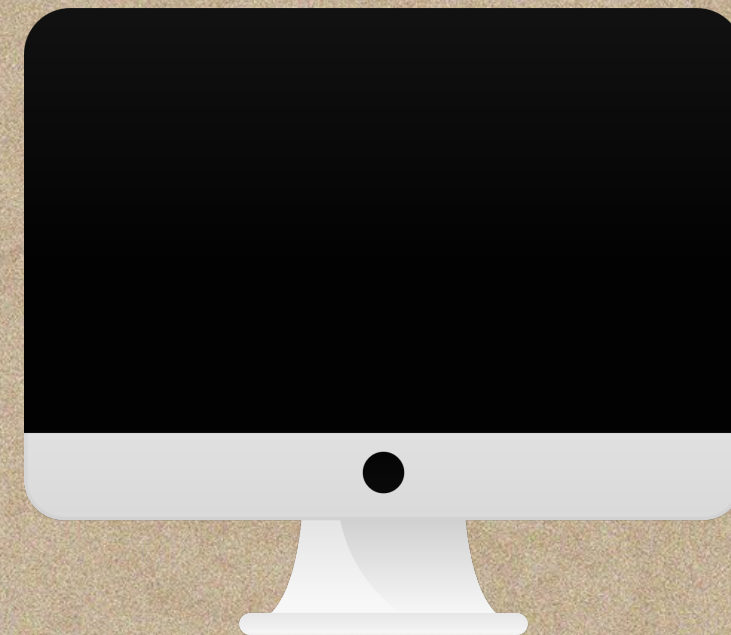
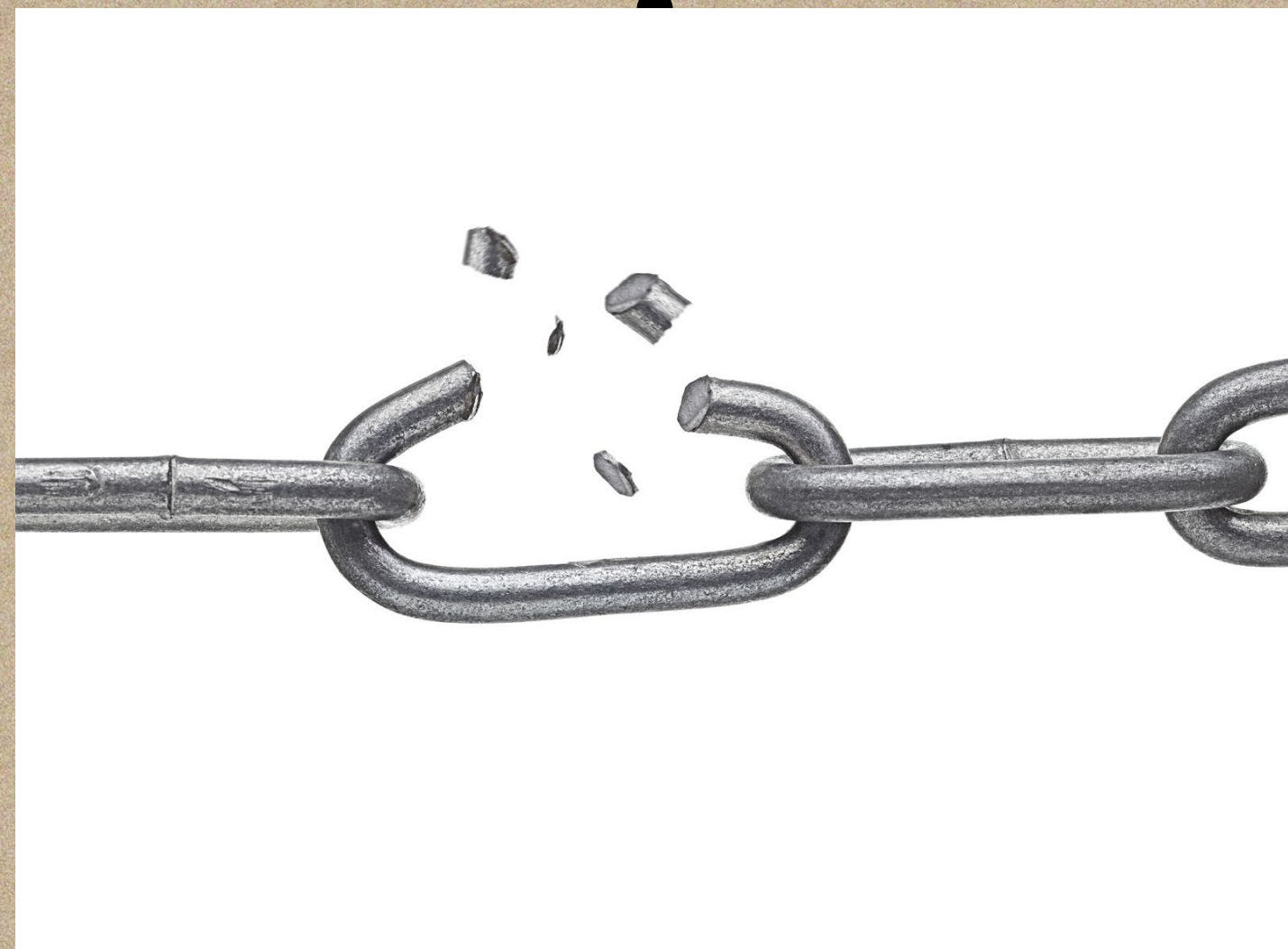
- iTunes library backups
- System files
- Full-length feature films



Rebuilding In Design

Adobe CS3 for Mac

- Fonts incompatible with most of our Windows machines
- Links, etc. formatted for Mac
- Happily, we are pack rats



2009
iMac

End Result

Panning for Gold

- Initial appraisal: 54,873 files across 1,902 directories
- 102 full PDF issues across eight years, 2003-2011
- Gaps, but a significant addition to the collection, and only known copy of any type in many instances

Coda:


More where that came from. Maybe.

Web Archiving —————• Newspapers in Kentucky


- UK Libraries Web Archives established in 2018
- Established workflow for Kentucky Kernel
(Libraries have collected since 1915)
- Workflow + manual in place, but Kentucky
Lantern was our first external newspaper
website accession

The Kentucky Lantern

HOME ABOUT SUBSCRIBE DONATE PART OF STATES NEWSROOM



ECONOMY ENVIRONMENT GOVERNMENT HEALTH ELECTION 2024



Please, don't shoot the wild pigs. It only makes them more elusive.

BY LIAM NIEMEYER - MAY 16, 2024

With GOP opposed, U.S. Senate panel advances bills to combat AI in elections

BY ZACHARY ROTH - MAY 16, 2024

U.S. Senate GOP tries to block states from spending some of their COVID relief cash

BY JENNIFER SHUTT - MAY 15, 2024

BRIEF

Kentucky K9 officer will sniff out evidence in child exploitation cases

BY SARAH LADD - MAY 15, 2024

U.S. Department of Justice investigating Kentucky juvenile detention conditions

BY SARAH LADD AND DEBORAH YETTER - MAY 15, 2024


CNN sets first Biden-Trump presidential debate for June 27 in Atlanta

BY ARIANA FIGUEROA - MAY 15, 2024


Live
Site

You are viewing an archived web page collected at the request of [University of Kentucky Libraries](#) using [Archive-It](#). This page was captured on 17:41:55 May 06, 2024, and is part of the [Kentucky Digital Newspaper Program Web Archives](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. Found 0 archived media items out of 0 total on this page.

HOME ABOUT SUBSCRIBE DONATE PART OF STATES NEWSROOM



ECONOMY ENVIRONMENT GOVERNMENT HEALTH ELECTION 2024



Plummeting balance in federal crime victims fund sparks alarm

Kentucky nursing-home industry says Biden staffing mandate 'impossible' to meet

BY MELISSA PATRICK, KENTUCKY HEALTH NEWS - MAY 6, 2024

Biden administration to issue rule expanding DACA health care access

BY ARIANA FIGUEROA - MAY 3, 2024

More Kentucky children have kept Medicaid coverage compared with other states

BY DEBORAH YETTER - MAY 3, 2024

Doctors plead with Congress to help improve U.S. maternal mortality rates

BY JENNIFER SHUTT - MAY 2, 2024

Archived

Metadata across platforms

Title: Kentucky Lantern website

URL: <https://kentuckylantern.com>

Description: The Kentucky Lantern is an independent, nonpartisan, free news service based in Frankfort, Kentucky, part of States Newsroom, a nonprofit network of journalists that works to fill gaps in state government reporting caused by the declining numbers of state and local journalists.

Captured 14 times between July 21, 2023 and May 06, 2024

Videos: 3 Videos Captured

Subject: Health -- Kentucky, Government -- Kentucky, Environment -- Kentucky, Economy -- Kentucky, Journalism -- Kentucky

Creator: Kentucky Lantern

Coverage: Kentucky

Format: website

Type: electronic newspaper

Rights: "Our stories may be republished online or in print under Creative Commons licence CC BY-NC-ND 4.0."

Collector: University of Kentucky Libraries Special Collections Research Center, University of Kentucky Libraries

Archive-It: Dublin Core

Metadata across platforms

Kentucky Lantern website Accession

Basic Information

Title	Kentucky Lantern website
Identifier	2023ms038
Accession Date	2023-08-17
Publish?	True
Content Description	The Kentucky Lantern is an independent, nonpartisan, free news service based in Frankfort, Kentucky, part of States Newsroom, a nonprofit network of journalists that works to fill gaps in state government reporting caused by the declining numbers of state and local journalists. Records harvested through Archive-It, crawled monthly.
Acquisition Type	Gift
Resource Type	Papers
Restrictions Apply?	False
Access Restrictions?	False
Use Restrictions?	False

Created by mcdonnell 2023-08-17 10:55:09 -0400 | Last Modified by mcdonnell 2023-10-06 10:22:57 -0400 | URI: /repositories/2/accessions/11

Dates

Creation	2023
Expression	2023
Type	Inclusive Dates
Label	Creation

Extents

ArchivesSpace:
DACs

Preparing for problems

ArchiveReady.com

 ArchiveReady
website archivability evaluation tool

[Help](#) [FAQ](#) [API](#) 

Is your website Archive Ready?

[Check now »](#)

What is ArchiveReady?

An online tool which evaluates **if a website will be archived correctly** by web archives, such as the Internet Archive.

Who is it for?

Web professionals who need to check if their websites are archive ready.

Web archive engineers who need to evaluate target websites before harvesting and ingestion.

Students who need to learn more about digital archiving.

Why bother?

Because **not all websites are archive ready** and this may result in invalid web archives and ultimately in **information loss**.

How does it work?

ArchiveReady analyses your website (i.e. HTML, Images, CSS, JS, Sitemaps) and performs complex evaluations in order to calculate **Website Archivability** from a set of Archivability Facets: Accessibility, Cohesion, Metadata & Standards Compliance.

Preparing for problems ---



Webrecorder
Web archiving for all!

 **ReproZip**
The Reproducibility Packer!

Takeaways

- Paper --> Microfilm --> Digitization
- Born-Digital--> Stay flexible.
- Publication participation is key, but sometimes active intervention is necessary.
- Advocacy is crucial. Benign neglect can't save us this time.

References

- About Us. La Voz, captured by Wayback Machine 11 November 2012, <https://web.archive.org/web/20111122000820/http://www.lavozky.com/sitertools.php?section=l&lang=eng>. Accessed 14 May 2024.
- Banos, Vangelis. Archiveready.com, <https://archiveready.com/>. Accessed 14 May 2024.
- Boss, Katy, Vicky Rampin, Rémi Rampin, and Ilya Kremer. Reprozip-Web. <https://web.reprozip.org>. Accessed 14 May 2024.
- Bryan, Ruth and Collier, Emily. Web Archiving Program Description Procedures. https://docs.google.com/document/d/1hF5hDegeZTU7D4s9hCMLVjGa1Or7_GGcCv-ZJfUnAAy. Accessed 14 May 2024.
- Collier, Emily. Web Archiving Program Manual. 2024, https://docs.google.com/document/d/1Z5ksBRlhYQ0BKDrsmWzm_ByEHT9G5H2u2HM6sGunVSs. Accessed 14 May 2024.
- Dappert, Angela & Jackson, Andrew & Kimura, Akiko. (2013). "Developing a Robust Migration Workflow for Preserving and Curating Hand-held Media". iPRES presentation, https://www.researchgate.net/publication/256762565_Developing_a_Robust_Migration_Workflow_for_Preserving_and_Curating_Hand-held_Media. Accessed 14 May 2024.
- Kentucky Kernel catalog record. InfoKat, University of Kentucky Libraries. https://saalck-uky.primo.exlibrisgroup.com/permalink/01SAA_UKY/15remem/alma9910867146802636. Accessed 14 May 2024.
- Kentucky Lantern. Launch Landing Page. captured via Wayback Machine on November 30, 2023. <https://web.archive.org/web/20221121014139/https://kentuckylantern.com/>. Accessed 14 May 2024.
- Library of Congress. "Recommended Format Statement: Web Archives". Library of Congress, <https://www.loc.gov/preservation/resources/rfs/webarchives.html>. Accessed 14 May 2024.
- Rothrock, Michelle, & Rhonemus, Alison, & Krabbenhoef, Nick. "Assessing High-volume Transfers from Optical Media at NYPL". Code4Lib Journal, 2017, <https://journal.code4lib.org/articles/15908>. Accessed 14 May 2024.
- Terry, Kopana, "The Digitization of Historic Newspapers on Microfilm: The Kentucky Experience", UK Knowledge, 2009, Library Faculty and Staff Publications. 256. https://uknowledge.uky.edu/libraries_facpub/256. Accessed 14 May 2024.
- Terry, Kopana and Weig, Eric. "Preserving Kentucky's Newspapers: Analogue Beginnings to Digital Frontier". Paper presented at: IFLA WLIC 2015 - Cape Town, South Africa in Session 80 - News Media. In: IFLA WLIC 2015 - Cape Town, South Africa. <https://library.ifla.org/id/eprint/1268>. Accessed 14 May 2024.
- University of Kentucky Web Archives. "Kentucky Digital Newspaper Web Archives". <https://archive-it.org/collections/21644>. Accessed 14 May 2024.
- van der Knijff, Johan. "Image and Rip Optical Media Like A Boss", Open Preservation Foundation", Open Preservation Foundation, 2017, <https://openpreservation.org/blogs/image-and-rip-optical-media-like-a-boss/>. Accessed 14 May 2024.
- Web Archiving at the University of Kentucky Libraries. University of Kentucky Libraries, https://libguides.uky.edu/web_archiving. Accessed 14 May 2024.



Andrew McDonnell
Digital Archivist
University of Kentucky
Special Collections Research Center
mcdonnell@uky.edu