

Preserving the Scholarly Web: Portable Web Archives for ETDs

Lauren Ko and Mark Phillips
University of North Texas Libraries
June 3, 2026

Web Archives in Libraries and Archives

Many of us are doing some sort of web archiving:

- Capturing the output of the university
- Extending physical collections into the digital environment
- Building born-digital collections

Most of us are using hosted tools for building these collections like:

- Archive-It
- Browsertrix

The image displays three overlapping screenshots of web archiving services. The top screenshot is the Browsertrix homepage, which includes a navigation bar with 'WEBRECORDER', 'Tools', 'Blog', 'Resources', 'About', 'Pricing', 'Sign In', and 'Sign Up'. The main heading reads 'Archive, Curate, and Share the Web.' Below this, it states 'High-fidelity crawling, indexing, and archiving empower you to preserve your digital content. Built and run by Webrecorder.' A 'Get Started!' button is visible. The middle screenshot is the Archive-It homepage, featuring the Archive-It logo and the text 'The leading service for preserving and accessing digital cultural heritage'. It also includes a 'Get Started!' button and a note 'Built at Internet Archive'. The bottom screenshot is the Internet Archive homepage, prominently displaying the 'Wayback Machine' search interface. It includes a search bar with the prompt 'Enter a URL or words related to a site's home page' and a 'SEARCH' button. Other elements include a 'DONATE' button, a 'Save Page Now' section, and a list of 'Tools' such as 'Wayback Machine Availability API', 'Chrome Extension', 'Firefox Add-on', 'Safari Extension', 'MS Edge Add-on', 'iOS app', and 'Android app'.

Built to Scale Up

- Create broad collections in crawling system
- Structure collections around subjects, entities, events
- Add seeds and set a crawl schedule
- WARC files are indexed with WARC6s from other crawls
- View crawl content organized by collection

The image displays three overlapping screenshots of the ARChive IT web interface. The top screenshot shows the main dashboard for the University of North Texas web archive, with a navigation bar and a 'Home' button. The middle screenshot shows the 'Texas Municipalities Web Archive' overview page, including a 'Current Subscription' section and a 'Collection Data' table. The bottom screenshot provides a detailed 'Crawl Overview' for a semiannual crawl, featuring a dashboard with key metrics and a donut chart.

| Metric | Value |
|----------------|-----------|
| Total Data | 178.8 GB |
| New Data | 176.1 GB |
| Total Docs | 1,208,754 |
| New Docs | 619,489 |
| Unchanged Docs | 589,265 |

#2638780
Collection: Texas Municipalities Web Archive
Archived: December 8, 2025 5:29 AM Central Daylight Time
Completed: December 15, 2025 7:08 AM Central Daylight Time
Crawling Technology: Standard

Crawl Frequency: Semiannual
Schedule (by IANA): Next Crawl: June 8, 2026 6:29 AM Central Daylight Time

Finished: Time Limit
Document Rate: /s
307.05 KB/s

178.8 GB Capture Data
New Data vs Duplicate Data: 2.9% (New Data), 97.0% (Duplicate Data)

Scoping Rules for this Crawl:
▶ Crawl Limits (3)
▶ Collection Scope Rules (0)
▶ Seed Scope Rules (1334)

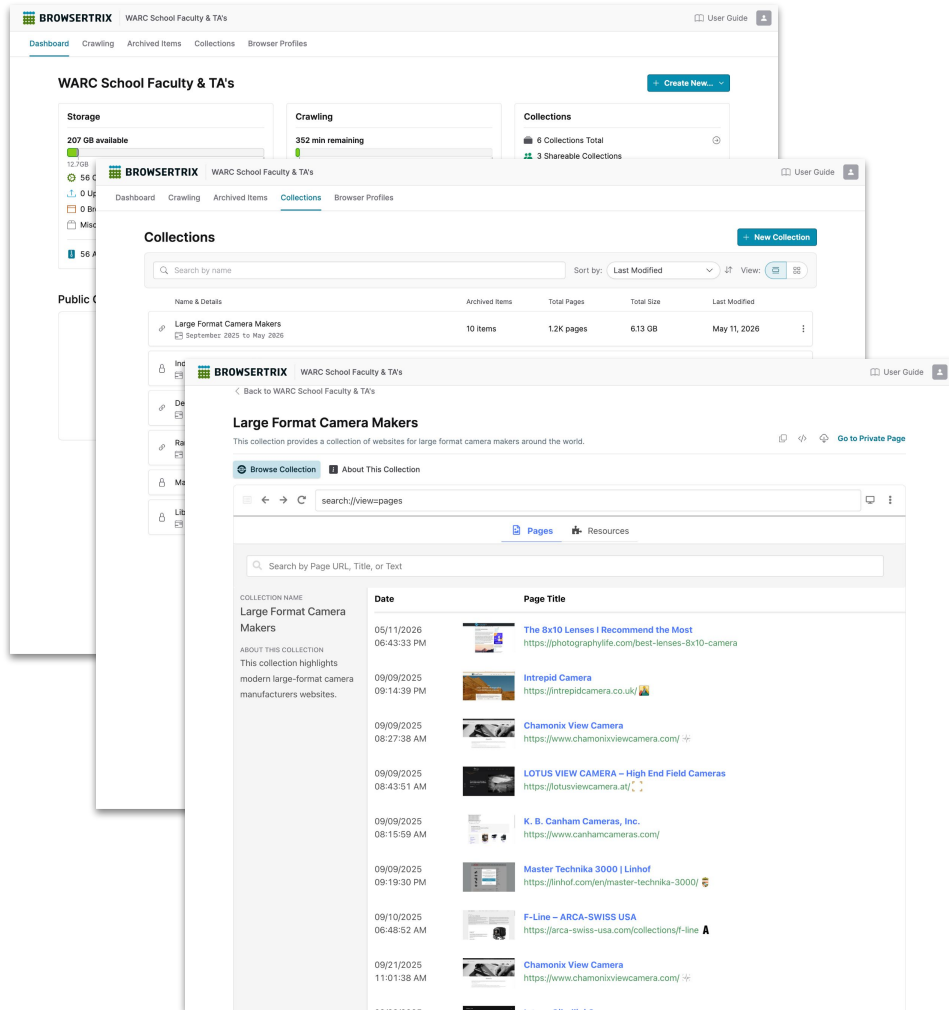
Notes (0)
Add Note

Scaling down is harder

Managing more collections/crawls adds overhead at crawling, processing, preservation, and access.

One time crawls of a website, part of a website, or individual pages/resources may not fit into a collection scope.

It takes work to extract part of the collection from a set of WARCs.



Web Archive Collection Zipped (WACZ)

- Newer standard for grouping and sharing WARC files and related material
- Zip file with a standardized file layout
- Playback tools designed to work with this format
- Makes a web archive collection portable
- Enables small web archives to exist by themselves

The image shows a GitHub repository for 'py-wacz' and a PDF document titled 'Web Archive Collection Zipped (WACZ)'. The GitHub repository page displays the repository name, a description, and a table of contents. The PDF document provides detailed information about the WACZ format, including its purpose, structure, and usage.

py-wacz public

Format Description Categories >> Browse Archived List

Web Archive Collection Zipped

Table of Contents

- Identification and description
- Local use
- Sustainability factors
- Quality and functionality factors
- File type identifiers
- Notes
- Format specifications
- Useful references

Identification and description

Web Archive Collection Zipped

Web Archive Collection Zipped (WACZ)

Webrecorder Recommendation 03 June 2021

More details about this document

This version:

Latest published version:

Use Cases for Decentralized Web Archives

Previous versions

Repository

Copyright © 2021 Webrecorder CC-BY

Abstract

WACZ is a media type that allows web archive collections to be packaged and shared on the web as a discrete file. A WACZ file includes all the data that is needed for the rendering archived content as well as contextual information required for users to interpret it. Rendering software can obtain this data on demand using HTTP Range requests, without requiring the entire file to be fully retrieved, or for it to be otherwise mediated by specialized server side software.

1. Conformance

As well as sections marked as non-normative, all authoring guidelines, diagrams, examples, and notes in this specification are non-normative. Everything else in this specification is normative.

The key words **MAY**, **MUST**, **MUST NOT**, and **SHOULD** in this document are to be interpreted as described in BCP.14 [RFC2119] [RFC8174] when, and only when, they appear in this specification, as shown here.

2. Status of This Document

WACZ Structure

WACZ files are zip files without compression.

This allows for them to be read in a similar way as WARC Files.

The *archive* folder contains WARC files.

The *indexes* folder contains a CDXJ index for resource lookup.

The *pages* folder holds Metadata about pages/resources in the WARC files.

datapackage files give an overview of the included files.

| Name | Date Modified | Size | Kind |
|------------------------------------------------------------------|-------------------------|-----------|-------------------|
| archive | Today at 6:38 AM | -- | Folder |
| UNT-b74b046de559-submission-43...025-20250610154615494-1.warc.gz | Jun 10, 2025 at 3:50 PM | 26.3 MB | gzip co...archive |
| UNT-b74b046de559-submission-43...025-20250610154615632-0.warc.gz | Jun 10, 2025 at 3:48 PM | 5.6 MB | gzip co...archive |
| UNT-b74b046de559-submission-43...shots-20250610154626065.warc.gz | Jun 10, 2025 at 3:48 PM | 40.6 MB | gzip co...archive |
| datapackage-digest.json | Jun 10, 2025 at 3:50 PM | 117 bytes | JSON |
| datapackage.json | Jun 10, 2025 at 3:50 PM | 2 KB | JSON |
| indexes | Today at 6:38 AM | -- | Folder |
| index.cdx.gz | Jun 10, 2025 at 3:50 PM | 56 KB | gzip co...archive |
| index.idx | Jun 10, 2025 at 3:50 PM | 479 bytes | Index file |
| logs | Today at 6:38 AM | -- | Folder |
| 20250610154612906.log | Jun 10, 2025 at 3:50 PM | 17 KB | Log File |
| pages | Today at 6:38 AM | -- | Folder |
| extraPages.jsonl | Jun 10, 2025 at 3:46 PM | 84 bytes | Document |
| pages.jsonl | Jun 10, 2025 at 3:50 PM | 3 KB | Document |

```
submission-4365-3889691948057319530_FISHER-PRIMARY-2025
├── archive
│   ├── UNT-b74b046de559-submission-4365-3889691948057319530_FISHER-PRIMARY-2025-20250610154615494-1.warc.gz
│   ├── UNT-b74b046de559-submission-4365-3889691948057319530_FISHER-PRIMARY-2025-20250610154615632-0.warc.gz
│   └── UNT-b74b046de559-submission-4365-3889691948057319530_FISHER-PRIMARY-2025-screenshots-20250610154626065.warc.gz
├── datapackage-digest.json
├── datapackage.json
├── indexes
│   ├── index.cdx.gz
│   └── index.idx
├── logs
│   └── 20250610154612906.log
├── pages
│   ├── extraPages.jsonl
│   └── pages.jsonl
└── 5 directories, 10 files
```

Case Study: UNT ETDs

Creating a Web Archive for each ETD

- Python script extracts URLs from ETD PDF
- Bash script loops through ETD URLs files and runs a crawl per each
- Crawls with Browsertrix Crawler running via Docker
- Produces a WACZ file we can package with the ETD

```
#!/ This is a script that loops through a set of ETD submissions' extracted URLs files
# and runs a browsertrix-crawl for each.

# Pass argument of list of .urls files via commandline, such as
# 2023-December/*/*.pdf.urls
for datafile in "$@"; do
    # Get absolute path of seed file.
    datafile=$(realpath "$datafile")
    # Get just the seed file name.
    filename=$(basename "$datafile")
    # Get the submission part of the path.
    submission=$(echo "$datafile" | awk -F/ '{print $(NF-1)}')
    # Get name of the PDF the URLs came from.
    pdfname="${filename%.*}"
    # Get the PDF name without an extension.
    rootname="${pdfname%.*}"
    # Set a name for the collection used by Browsertrix Crawler.
    collectionname="${submission}_${rootname}"
    # Replace any spaces with underscore.
    collectionname="${collectionname// /_}"
    echo "Processing $filename to $collectionname"
    docker run -p 8061:8061 -v "$datafile":/app/seedFile.txt -v $PWD/crawls:/crawls/ webrecorder/
    browsertrix-crawler crawl --seedFile /app/seedFile.txt --generateWACZ --scopeType page --generate
    CDX --screenshot thumbnail,fullPage --screenshotPort 8061 --diskUtilization 95 --description "Data
    served from URLs parsed from ETD $pdfname" --warcinfo.operator "University of North Texas Librar
    ies" --collection "$collectionname" --delay 6 --warcPrefix "UNT" --workers=2
    cp "$datafile" "$PWD/crawls/collections/$collectionname/"
done
```

Issues encountered with Crawling

Authors introduced typos in DOI URLs that caused Browertrix Crawler to abort crawls.

- <http://doi.10.1080/19315860802611415> (missing .org/; domain read as .1080)
- We aren't fixing all author typos in ETDs, but Tessa at Webrecorder fixed our issue of [aborted crawls on a failed seed](#)

Sites DOIs commonly point to may use bot-detecting services like Cloudflare's making them difficult to archive, even with browser-based crawling.

Including multiple speculative versions of seeds increases 404 noise in WACZ.

- Making live requests with Python to check URLs for 200 response at text extraction time (to prevent multiple versions) were often identified by the server as a bot traffic and affected the response.

Organization of an ETD Submission Package

- submission_3270
 - 01_jpg/
 - 000100tp.jpg
 - 00020000.jpg
 - 0003r001.jpg
 - 02_pdf/
 - ARIAS-THESIS-2023.pdf
 - 03_vireo/ (*preservation only*)
 - License.txt
 - Permissions.pdf
 - 04_wacz/
 - ARIAS-THESIS-2023.pdf.urls
 - submission_3270_ARIAS-THESIS-2023.wacz

University Libraries / UNT Digital Library / **Results** / About / All Formats

Discriminative Control of Behavioral Variability in Video Game Play **(All Formats)**

This item is available in multiple formats. Choose one below.

[Read this thesis online.](#)

| | |
|------------------|-------------|
| Number of items: | 50 |
| Filetype: | jpg (image) |

[Download this thesis.](#)

| | |
|------------------|------------|
| Number of items: | 1 |
| Filetype: | pdf (file) |

[Download this thesis.](#)

| | |
|------------------|-------------|
| Number of items: | 2 |
| Filetype: | wacz (file) |

WACZ in Repository

Now that we have WACZ files in the repository, we have integrated playback in the UNT Digital Library.

We let users know there are archived web resources available for an ETD.

Any item in the digital library with a WACZ includes a link to "archived web resources".

<https://digital.library.unt.edu/ark:/67531/metadc2179336/>

The screenshot displays the UNT Digital Library interface. The top navigation bar includes the UNT logo and links for HOME, COLLECTIONS, PARTNERS, TITLES, LOCATIONS, TYPES, DATES, ABOUT, TOUR, and CONTACT US. The main content area is divided into a left sidebar and a main content pane. The sidebar contains sections for 'About This Dissertation' (Overview, Who, What, When), 'Search Inside' (Search Inside), 'Read Now' (Start Reading, Magnify First Page, Jump to..., Show All Pages, All Formats), 'Print & Share' (social media icons), and 'Citations, Rights, Re-Use' (Citing This Dissertation, Responsibilities of Use, Licensing & Permissions, Linking & Embedding, Copies & Reproductions). The main content pane features the title 'The University for Who? Student Narratives of Native Identity, Belonging, and Navigating a Racialized Organization' by Emilia Morgan Gaston, MS, MA. It includes a description of the dissertation, a list of approvers (William Scarborough, Ming Professor; Craig Morgan, Graduate Member; Phillip W. Carr, Member; Matthew P. Coleman, Member; Dawn B. G. ...), and a 'Description' section. The description states: 'This qualitative case study aims to understand the ways in which students identifying as Native American, American Indian, and Indigenous navigate attending a university informed by their identities. Through semi-structured interviews with Indigenous students and participant observation with a Native American student organization, this study identified how this demographic of students navigate and conceptualize their identities as Native and Indigenous peoples, the benefits of joining a Native American student organization on their university campus, and how they experience the university as a racialized organization. One overarching and three nuanced research questions were examined to illustrate how students' identities inform how ... continued below'. Below the description is a 'Description' section with the text: 'This dissertation is part of the collection entitled: UNT Theses and Dissertations and was provided by the UNT Libraries to the UNT Digital Library, a digital repository hosted by the UNT Libraries. It has been viewed 150 times, with 69 in the last month. More information about this dissertation can be viewed below.' At the bottom of the main content pane, there is a search bar and a green button labeled 'Search'. Below the search bar, there are two icons: a PDF icon with the text 'PDF Version Also Available for Download.' and a globe icon with the text 'This item contains archived web resources.'

WACZ Playback

Users can download WACZ and use a service like <https://replayweb.page/> to playback content.

- Replay from local WACZ copy
- Replay via link to UNT-hosted WACZ

UNT Digital Library includes a local replayweb.page integration.

The screenshot shows the UNT Digital Library interface. The top navigation bar includes 'HOME', 'COLLECTIONS', 'PARTNERS', 'TITLES', 'LOCATIONS', 'TYPES', 'DATES', 'ABOUT', 'TOUR', and 'CONTACT US'. The main content area displays a search result for 'Village Creek, Battle of' with a URL: <https://www.tshaonline.org/handbook/entries/village-creek-battle-of>. The result is dated 9/3/2023, 10:33:24 AM. The page content includes a title 'Village Creek, Battle of' by Donald Fraxier and a detailed description of the battle. The description states: 'VILLAGE CREEK, BATTLE OF: The battle of Village Creek, fought on May 24, 1841, was a running gunfight along the banks of Village Creek, a major tributary of the Trinity River, in eastern Tarrant County. The stream now forms the city limits of Arlington and Fort Worth, and much of the battlefield has been inundated by Lake Arlington. A series of American Indian villages situated along the creek, housing Caddos, Cherokees, and Tonkawas, served as a stronghold against the encroachments of White settlers from the east and Comanches from the west. As frictions increased between settlers and Native Americans, the government of the Republic of Texas authorized a number of punitive raids against the Village Creek settlements. Two such expeditions launched in 1838 failed to locate the towns but did cause the American Indian groups to intensify their raids of frontier settlements. In 1841 Gen. Edward H.

| Filename | Size | Format | Download |
|-----------------------------------------------|----------|-----------------|--------------------------|
| GASTON-DISSERTATION-2023.pdf.urls | 7.5 KB | text/plain | Download |
| submission_3542_GASTON-DISSERTATION-2023.wacz | 913.2 MB | application/zip | Download |

<https://digital.library.unt.edu/ark:/67531/metadc2179336/m4/>



About This Dissertation

- Overview
- Who
- What
- When

Search Inside

Search Inside

Read Now

- Start Reading
- Magnify First Page
- Jump to...
- Show All Pages 198
- All Formats 3

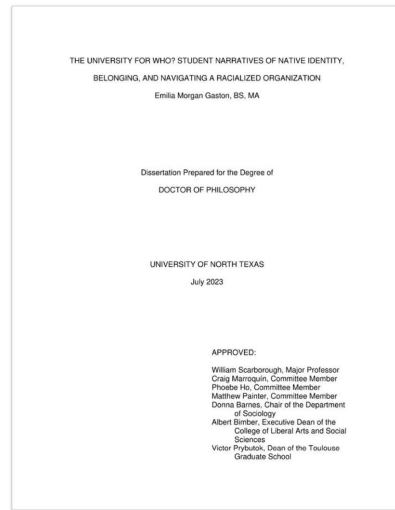
Print & Share



Citations, Rights, Re-Use

- Citing This Dissertation
- Responsibilities of Use
- Licensing & Permissions
- Linking & Embedding
- Copies & Reproductions

The University for Who? Student Narratives of Native Identity, Belonging, and Navigating a Racialized Organization



Showing 1-4 of 198 pages in this dissertation.

PDF Version Also Available for Download.

This item contains archived web resources.

Description

This qualitative case study aims to understand the ways in which students identifying as Native American, American Indian, and Indigenous navigate attending a university informed by their identities. Through semi-structured interviews with Indigenous students and participant observation with a Native American student organization, this study identified how this demographic of students navigate and conceptualize their identities as Native and Indigenous peoples, the benefits of joining a Native American student organization on their university campus, and how they experience the university as a racialized organization. One overarching and three nuanced research questions were examined to illustrate how students' identities inform how ... [continue below](#)

Creation Information

[Gaston, Emilia Morgan](#) July 2023.

Context

This **dissertation** is part of the collection entitled: [UNT Theses and Dissertations](#) and was provided by the [UNT Libraries](#) to the [UNT Digital Library](#), a digital repository hosted by the [UNT Libraries](#). It has been viewed 236 times, with 6 in the last month. More information about this dissertation can be viewed below.

Search



About This Dissertation

- Overview
- Who
- What
- When

Search Inside

Read Now

- Show All Pages 2
- All Formats 3

Print & Share



Citations, Rights, Re-Use

- Citing This Dissertation
- Responsibilities of Use
- Licensing & Permissions
- Linking & Embedding
- Copies & Reproductions

The University for Who? Student Narratives of Native Identity, Belonging, and Navigating a Racialized Organization (Archived Web Resources).

← → ↻ 📄 search://view=pages ⋮



🔍 Search by Page URL, Title or Text

| SUBMISSION_3542 | Date▲ | Page Title |
|--------------------------------------------------------|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| _GASTON- DISSERTATION- 2023.WACZ 94 Pages | 9/3/2023 9:33:24 AM | Village Creek, Battle of https://www.tshaonline.org/handbook/entries/village-creek-battle-of |
| | 9/3/2023 9:33:00 AM | Milestone Documents National Archives https://www.ourdocuments.gov/doc.php?flash=true&doc=50&page=transcript12 |
| | 9/3/2023 9:32:45 AM | Milestone Documents National Archives https://www.ourdocuments.gov/doc.php?flash=true&doc=50&page=transcript |
| | 9/3/2023 9:32:28 AM | Paying to Play Indian: The Dawes Rolls and the Legacy of \$5 Indians - ICT News https://newsmaven.io/indiancountrytoday/archive/paying-to-play-indian-the-dawes-rolls-and-the-legacy-of-5-indians-3yha0LIdYUaH7smRsrks8A/ |
| | 9/3/2023 9:31:52 AM | COE - College Enrollment Rates https://nces.ed.gov/programs/coe/indicator/cpb/college-enrollment- |

| Filename | Size | Format | Download |
|-----------------------------------------------|----------|-----------------|--------------------------|
| GASTON-DISSERTATION-2023.pdf.urls | 7.5 KB | text/plain | Download |
| submission_3542_GASTON-DISSERTATION-2023.wacz | 913.2 MB | application/zip | Download |

9/3/2023

9:32:28 AM

Paying to Play Indian: The Dawes Rolls and the Legacy of \$5 Indians - ICT News

<https://newsmaven.io/indiancountrytoday/archive/paying-to-play-indian-the-dawes-rolls-and-the-legacy-of-5-indians-3yha0LldYUaH7smRsrks8A/>



About This Dissertation

- Overview
- Who
- What
- When

Search Inside

Read Now

- Show All Pages 2
- All Formats 3

Print & Share



Citations, Rights, Re-Use

- Citing This Dissertation
- Responsibilities of Use
- Licensing & Permissions
- Linking & Embedding
- Copies & Reproductions

The University for Who? Student Narratives of Native Identity, Belonging, and Navigating a Racialized Organization (Archived Web Resources).

← → ↻ 📄 <https://ictnews.org/archive/paying-play-indian-dawes-rolls-legacy-5-indian> 9/3/2023, 9:32:29 AM ⋮



☰ NEWS NEWSCASTS CLASSIFIEDS CAREERS OUTSIDE PRESS POOL

🔍 DONATE

Paying to Play Indian: The Dawes Rolls and the Legacy of \$5 Indians

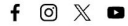
Paying a commissioner \$5 could get an opportunistic white person on the Dawes Rolls 125 years ago, which still causes problems for tribes today.

ALYSA LANDRY · UPDATED: SEP 13, 2018 · ORIGINAL: MAR 21, 2017



It may be fashionable to play Indian now, but it was also trendy 125 years ago when people paid \$5 apiece for falsified documents declaring them Native on the Dawes Rolls.

| Filename | Size | Format | Download |
|-----------------------------------------------|----------|-----------------|--------------------------|
| GASTON-DISSERTATION-2023.pdf.urls | 7.5 KB | text/plain | Download |
| submission_3542_GASTON-DISSERTATION-2023.wacz | 913.2 MB | application/zip | Download |



Paying to Play Indian: The Dawes Rolls and the Legacy of \$5 Indians

Paying a commissioner \$5 could get an opportunistic white person on the Dawes Rolls 125 years ago, which still causes problems for tribes today.

ALYSA LANDRY · UPDATED: SEP 13, 2018 · ORIGINAL: MAR 21, 2017



It may be fashionable to play Indian now, but it was also trendy 125 years ago when people paid \$5 apiece for falsified documents declaring them Native on the Dawes Rolls.

These so-called five-dollar Indians paid government agents under the table in order to reap the benefits that came with having Indian blood. Mainly white men with an appetite for land, five-dollar Indians paid to register on the Dawes Rolls, earning fraudulent enrollment in tribes along with benefits inherited by generations to come.


“These were opportunistic white men who wanted access to land or food rations,” said Gregory Smithers, associate professor of history at Virginia Commonwealth University

| Filename | Size | Format | Download |
|-----------------------------------------------|-------------|-----------------|--------------------------|
| GASTON-DISSERTATION-2023.pdf.urls | 7.5 KB | text/plain | Download |
| submission_3542_GASTON-DISSERTATION-2023.wacz | 913.2 MB | application/zip | Download |

<https://ssrn.com/abstract=2009877>
<https://podcasts.apple.com/us/podcast/the-ancestors-know-you-real-life-reconnection-stories/id1454424563?i=1000545327677identity>
<https://northtexan.unt.edu/issues/2020-spring/early-diversity>
<https://doi.org/10.1177/2332858417751896>
<https://scholarworks.uvm.edu/tvc/vol32/iss1/4>
<https://news.unt.edu/news-releases/unt-receives-national-diversity-award-fourth-year-row>
<https://nces.ed.gov/programs/coe/indicator/cpb/college-enrollment-ratecultural>
<https://govtrack.us.s3.amazonaws.com/legislink/pdf/stat/9/STATUTE-9-Pg574a.pdf>
<https://illinoisnewsroom.org/a-retired-university-of-illinois-race-based-mascot-still-lingers-on-campus/>
<https://extension.usu.edu/news/purpose-and-benefit-of-land-grant-extension-universities>
<https://podcasts.apple.com/us/podcast/the-ancestors-know-you-real-life-reconnection-stories/id1454424563?i=1000545327677>
<https://doi.org/10.2307/1354221>
<https://www.washingtonpost.com/outlook/2019/11/08/dark-history-land-grant-universities/>
<https://doi.org/10.2307/1341787>
<https://www.allmyrelationspodcast.com/post/ep-10-beyond-blood-quantum>
<https://www.globenewswire.com/news-release/2022/04/13/2421920/0/en/DNA-Test-Kits-Market-to-Cross-US-2-7-Bn-Mark-by-2030-TMR-Report.html>
<https://www.cherokee.org/all-services/tribal-registration/frequently-asked-questions/>
<https://blog.23andme.com/ancestry-reports/native-american-dna/>
https://institutionalresearch.unt.edu/sites/default/files/xf-1-new2020-2021-_unt.pdf
https://nces.ed.gov/programs/digest/d18/tables/dt18_302.60.asp
<http://dx.doi.org/10.18251/ijme.v10i2.137>
<https://www.muse.jhu.edu/article/745704>
<https://digitalcommons.law.seattleu.edu/ailj/vol10/iss2/4>
<https://texashistory.unt.edu/ark:/67531/metaph257183/#:~:text=Denton%20Independent%20School%20District,->
<https://doi.org/10.1002/ejsp.1972>
<https://www.tcu.edu/native-american-indigenous-peoples/index.php>
<https://www.apa.org/pi/oema/resources/indian-mascots#:~:text=The%20APA%20is%20calling%20upon,colleges%2C%20universities%2C%20athletic%20teams%20andtheir>
<https://nces.ed.gov/programs/coe/indicator/cpb/college-enrollment-rate>
<https://sites.ed.gov/whiaiane/tribes-tcus/tribal-colleges-and-universities/>
<https://doi.org/10.1080/10705422.2012.732001>
<https://files.eric.ed.gov/fulltext/ED527093.pdf>
<https://doi.org/10.1177/0042085915602542>
<https://nces.ed.gov/collegenavigator/?q=north+texas&s=all&id=227216#enrolmt>
<https://tinyurl.com/UNTNative>
https://www.tahlequahdailypress.com/news/updated-coalition-wants-misleading-statue-removed-at-nsu/article_181b8c9a-747a-567f-ad97-f493e6cf1c2f.html
<https://www.merriam-webster.com/dictionary/safe%20space>
<https://www.muse.jhu.edu/article/798557>
<https://www.tshaonline.org/handbook/entries/denton-tx-denton-county#:~:text=The%20city%20was%20not%20an,Denton>
<https://doi.org/10.1353/csd.2007.0039>
<https://www.washingtonpost.com/outlook/2019/11/08/dark-history-land-grant->
<https://www.technologyreview.com/s/610233/2017-was-the-year-consumer-dna-testing-blew-up>
<https://www.unt.edu/story/unt-clearly-university-north-texas#:~:text=Darton>
<https://www.edweek.org/leadership/more-states-push-schools-to-drop-native-american-mascots/2022/11>
<https://doi.org/10.1177/0013916518756988>
<https://www.archives.gov/milestone-documents/fort-laramie-treaty>
<https://traditions.unt.edu/history/history-unt>
<https://doi.org/10.1177/1177180118822833>
<https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html>
<https://uitct.org/mission-history/>
<http://www.jstor.org/stable/1185951>
<https://doi.org/10.1002/9780470999103.ch5>
<https://doi.org/10.1177/0748175610384811>

Explore and Replay Interactive Archived Webpages Directly in your Browser. [\(See Examples\)](#)

Load Web Archive

 Choose File...

Enter a URL or click 'Choose File' to select a WARC, WACZ, or HAR file

Load

 Hide

Loaded Archives

Don't have any web archives yet? Check out [ArchiveWeb.page](#) to save pages as you browse the web, or [sign up for Browsertrix](#) to archive entire websites with automated crawling!

Explore and Replay Interactive Archived Webpages Directly in your Browser. ([See Examples](#))

Load Web Archive

Choose File...

Load

Hide

Loaded Archives

Don't have any web archives yet? Check out [ArchiveWeb.page](#) to save pages as you browse the web, or [sign up for Browsertrix](#) to archive entire websites with automated crawling!



search://view=pages



Pages

Resources

Search by Page URL, Title, or Text

SUBMISSION_3542_GASTON
-DISSERTATION-2023.WACZ

94 Pages

Date▲

Page Title

9/3/2023
9:33:24 AM



Village Creek, Battle of
<https://www.tshaonline.org/handbook/entries/village-creek-battle-of>

9/3/2023
9:33:00 AM



Milestone Documents | National Archives
<https://www.ourdocuments.gov/doc.php?flash=true&doc=50&page=transcript12>

9/3/2023
9:32:45 AM



Milestone Documents | National Archives
<https://www.ourdocuments.gov/doc.php?flash=true&doc=50&page=transcript>

9/3/2023
9:32:28 AM



Paying to Play Indian: The Dawes Rolls and the Legacy of 5\$ Indians - ICT News
<https://newsmaven.io/indiancountrytoday/archive/paying-to-play-indian-the-dawes-rolls-and-the-legacy-of-5-indians-3yha0LLdYUaH7smRsrks8A/>

9/3/2023
9:31:52 AM



COE - College Enrollment Rates
<https://nces.ed.gov/programs/coe/indicator/cpb/college-enrollment->

9/3/2023
9:31:16 AM



Ethnicity without groups | European Journal of Sociology / Archives Européennes de Sociologie | Cambridge Core

Annual Reports and Information Staff (Annual Reports)



- Home
- About
- Condition of Education
- Digest of Education Statistics
- Projections
- Topical Studies
- Publications

Collapse

Family Characteristics

Preprimary, Elementary, and Secondary Education

Postsecondary Education

Postsecondary Students (8)

Immediate College Enrollment Rate

College Enrollment Rates

Impact of the Coronavirus Pandemic on Fall Plans for Postsecondary Education

Undergraduate Enrollment

Postbaccalaureate Enrollment

Characteristics of Postsecondary Students

College Student Employment

Postsecondary Enrollment Status of Youth From Rural Areas

Institutions (1)



Postsecondary Education

College Enrollment Rates

Last Updated: May 2023 | Suggested citation

Download PDF

The overall college enrollment rate for 18- to 24-year-olds decreased from 41 percent in 2010 to 38 percent in 2021. The college enrollment rate in 2021 was higher for 18- to 24-year-olds who were Asian (60 percent) than for those who were White (38 percent), Black (37 percent), of Two or more races (35 percent), Hispanic (33 percent), and American Indian/Alaska Native (28 percent).

The overall college enrollment rate of 18- to 24-year-olds (ages in which students traditionally enroll in college) was 38 percent in 2021. In this indicator, *college enrollment rate* is defined as the percentage of 18- to 24-year-olds enrolled as undergraduate or graduate students in 2- or 4-year institutions. The [Immediate College Enrollment Rate](#) indicator, in contrast, presents data on the percentage of high school completers who enroll in 2- or 4-year institutions in the fall immediately following high school.

Select a subgroup characteristic from the drop-down menu below to view relevant text and figures.

Closing

WACZ is a way of packing and transporting web archives that has the potential to change how we think about collecting material from the web.

There are many opportunities for “small web archives” to become more common in our collections.

Small web archives provide new avenues of preserving the scholarly record.

The screenshot shows the 'Web Archive Collection Zipped (WACZ)' specification document. On the left is a 'TABLE OF CONTENTS' with sections: Abstract, 1. Conformance, 2. Status of This Document, 3. Terminology, 4. Introduction (4.1 Motivation, 4.2 Existing Tools), 5. WACZ Object (5.1 Directory Layout, 5.2 Directories and Files (5.2.1 archive, 5.2.2 indexes, 5.2.3 pages.json, 5.2.4 datapackage.json, 5.2.5 datapackage-digest.json), 5.3 Other files and directories, 5.4 Zip Format (5.4.1 Zip Compression, 5.4.2 Zip Format File Extension), 6. Processing Model, 7. Publishing (7.1 Content-Length, 7.2 Partial Requests, 7.3 CORS, 7.4 Media Type, 7.5 Example Response), A. References (A.1 Normative references). On the right, the document title is 'Web Archive Collection Zipped (WACZ)' with a logo and a 'ReSpec' button. Below the title is 'Webrecorder Recommendation 03 June 2021' and a link for 'More details about this document'. It lists 'This version' and 'Latest published version' with URLs, 'Editors' (Ilya Kreymer and Ed Summers), 'Additional Documents' (Use Cases for Decentralized Web Archives), 'Previous versions' (1.1.0, 1.0.0), and 'Repository' (GitHub, Issues, Commits). A copyright notice '© 2021 Webrecorder CC-BY' is present. The 'Abstract' section explains that WACZ is a media type for packaging and sharing web archive collections, including data for rendering and contextual information. It notes that sections marked as non-normative (like 1. Conformance) contain guidelines, diagrams, and examples, while normative sections contain key words (MAY, MUST, MUST NOT, SHOULD) to be interpreted as described in BCP 14. The '1. Conformance' section is partially visible at the bottom.

Questions / Comments

Feel free to contact us if you have questions.

Lauren Ko – lauren.ko@unt.edu

Mark Phillips – mark.phillips@unt.edu