

Collecting Diachronic Affiliation Data for Faculty at HBCUs Using Memento

Deanna Zarrillo^{*}, Mat Kelly^{*}, Christopher Jackson[†], Erjia Yan^{*}

{dz364, mkelly, ey86}@drexel.edu — cjacks75@vols.utk.edu

^{*} Drexel University College of Computing and Informatics, Philadelphia, PA

[†] University of Tennessee - Knoxville, Knoxville, TN

Project Overview & Agenda

Key points...

- Supports the efforts of a larger project on brain drain
- Large-scale longitudinal HBCU faculty dataset
- Associates live links with past versions

What I'll cover...

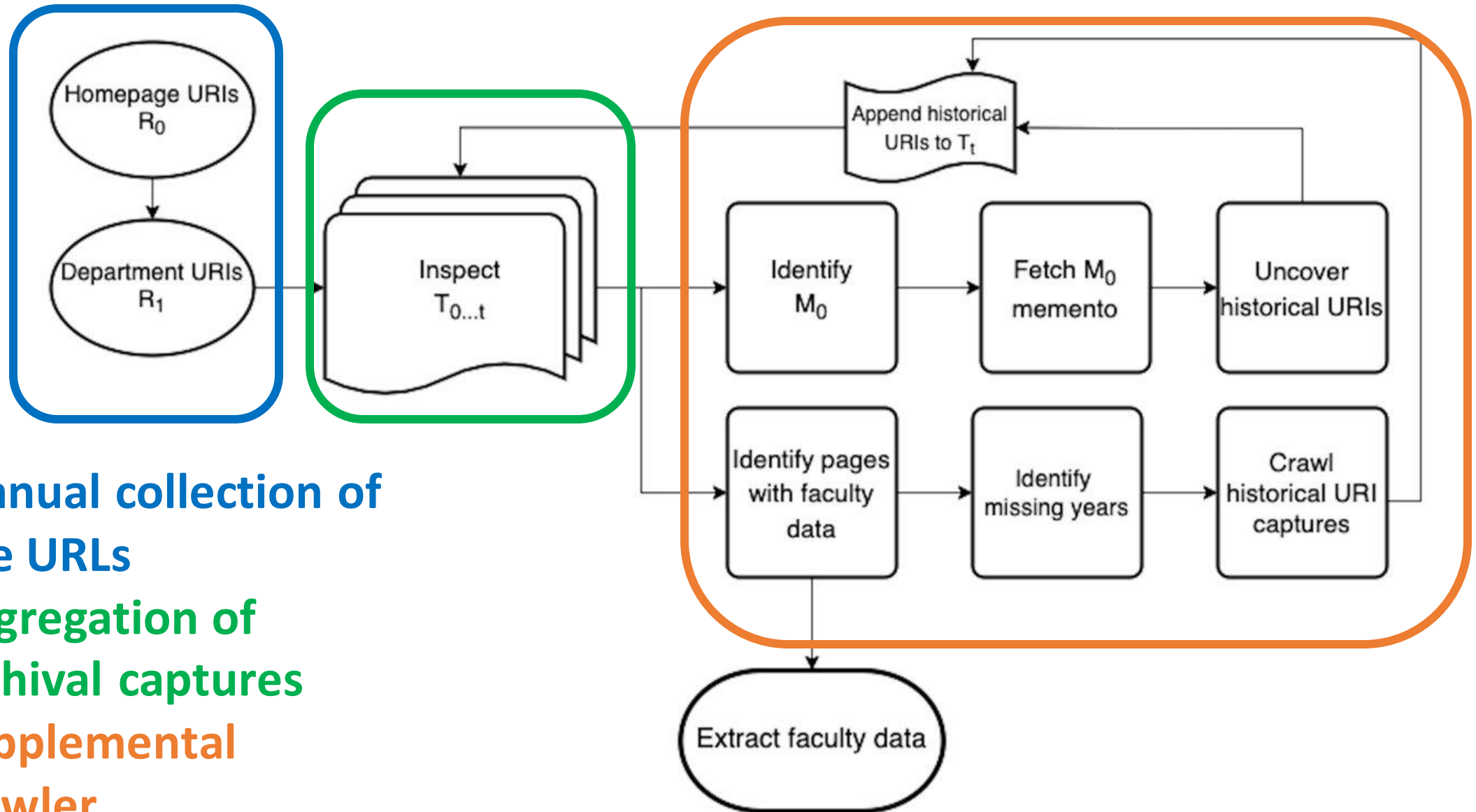


Background & Definitions

- Historically Black Colleges and Universities (HBCUs)
 - 3% of higher education institutions
 - 80% Black federal judges
 - 85% Black doctors
 - 75% Black PhD graduates, etc.
- "Brain drain"
 - Emigration of highly trained people from a particular country
 - **Hypothesis:** Emigration of faculty from HBCUs --> elsewhere
 - Goal: uncover motivations for movement
- Need for large scale analysis
- Need for data!



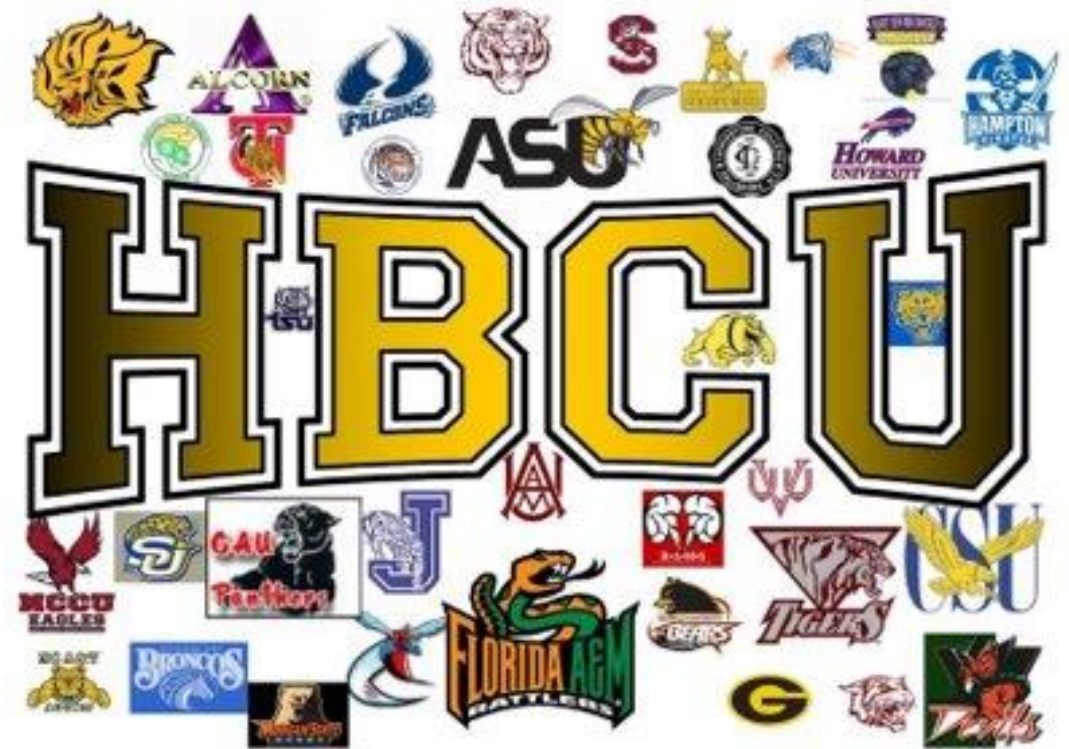
Workflow Overview



1. Manual collection of live URLs
2. Aggregation of archival captures
3. Supplemental crawler

Step 1: Manual Collection

- University Websites
- 35 HBCUs
 - 11 doctoral
 - 24 masters
- Estimated 10k faculty according to Carnegie Classification of Institutions of Higher Education (CCIHE)
- Utilize several web archives (2006-2021)
 - Collection of homepage and department level page URIs for each HBCU



Step 2: Aggregation and Association

- MemGator
- Memento aggregator queries multiple archives
- TimeMap response – file of aggregated archival URIs (URI-Ms) and their relative relations
- Obstacle: departmental coverage in archives
- Solution: supplemental web crawler

```
<http://howard.edu>; rel="original",  
<http://aggregator.example/timemap/link/http://howard.edu>; rel="self"; type="application/link-format",  
<http://web.archive.bibalex.org:80/web/19971212074755/http://www.howard.edu/>; rel="first memento";  
datetime="Fri, 12 Dec 1997 07:47:55 GMT",  
<https://wayback.archive-it.org/all/19971212074755/http://www.howard.edu:80/>; rel="memento";  
datetime="Fri, 12 Dec 1997 07:47:55 GMT",  
<https://web.archive.org/web/19971212074755/http://www.howard.edu:80/>; rel="memento"; datetime="Fri,  
12 Dec 1997 07:47:55 GMT",  
...  
<http://web.archive.bibalex.org:80/web/20000302015944/http://www.howard.edu/>; rel="memento";  
datetime="Thu, 02 Mar 2000 01:59:44 GMT",  
...  
<https://web.archive.org/web/20220410092256/https://howard.edu/>; rel="memento"; datetime="Sun, 10 Apr  
2022 09:22:56 GMT",  
<https://web.archive.org/web/20220410212257/https://howard.edu/>; rel="last memento"; datetime="Sun, 10  
Apr 2022 21:22:57 GMT",  
<http://aggregator.example/timemap/link/http://howard.edu>; rel="timemap"; type="application/link-format",  
<http://aggregator.example/timemap/json/http://howard.edu>; rel="timemap"; type="application/json",  
<http://aggregator.example/timemap/cdxj/http://howard.edu>; rel="timemap"; type="application/cdxj+ors",  
<http://aggregator.example/timegate/http://howard.edu>; rel="timegate"
```

Original URI (URI-R)

Relative Relations

memento (URI-M)

Other TimeMaps
(URI-Ts)

Step 3: Crawling for Supplemental Data

- Custom crawler:
 - Web archive domain specificity
 - Workflow integration
- Seed frontier approach with priority buckets
 - Current work: model development
 - Upcoming work: association via quasi-canonicalization

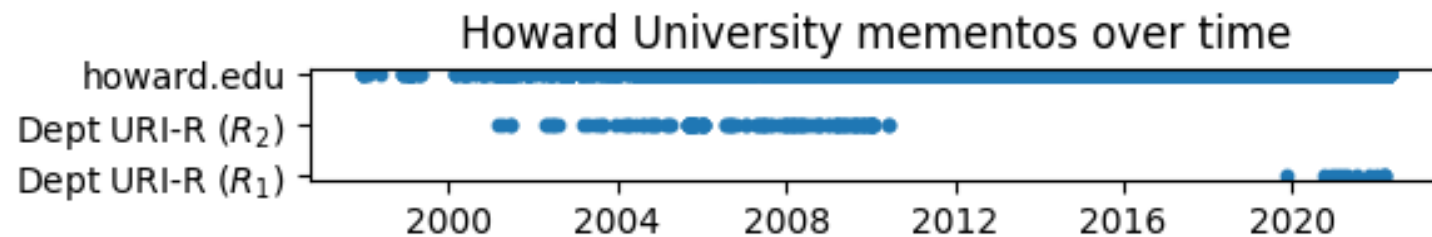


Image from: Kelly, M., Zarrillo, D., Jackson, C., & Yan, E. (2022). *First steps in identifying academic migration using memento and quasi-canonicalization*. Web Archiving and Digital Libraries Workshop 2022 at the ACM/IEEE Joint Conference on Digital Libraries.

Future Work & Implications

Refine and
finalize
process/data
collection

Clean,
supplement,
classify, connect

Surveys and
interviews

Ethical considerations of archival data use (Maemura, 2018)

Acknowledgements *Thank you to our collaborators!*

Research support by NSF SoS:DCI #2122525



Summary:

1. Hypothesis: HBCUs experience brain drain
2. Empirical studies on HBCU faculty mobility have not been conducted since the 90s.
3. We use manual data collection, Memento aggregator, and a custom web crawler to facilitate the creation of a large-scale longitudinal dataset.
4. This dataset facilitates investigation into the mobility motivations for HBCU faculty.

- Allen, H. L. (1991). The Mobility of Black Collegiate Faculty Revisited: Whatever Happened to the " Brain Drain"? *The Journal of Negro Education*, 60(1), 97-109.
- Gasman, M. (2013). The changing face of historically Black colleges and universities.
- Seymore, S. B. (2005). I'm confused: How can the federal government promote diversity in higher education yet continue to strengthen historically Black colleges. *Wash. & Lee J. Civil Rts. & Soc. Just.*, 12, 287.
- NCES. (2020). College navigator. Retrieved from <https://nces.ed.gov/COLLEGENAVIGATOR/?s=all&sp=4&pg=1>
- Minor, J. T. (2004). Introduction: Decision making in historically Black colleges and universities: Defining the governance context. *Journal of Negro Education*, 40-52.
- Walters, A. (2005). Predominantly Black and historically Black colleges spar over federal funds. *The Chronicle of Higher Education*, 52(6), A28.
- Rights, U. C. o. C. (2010). The educational effectiveness of historically Black colleges and universities: Author Washington, DC.
- Alam, S. & Nelson, M. L. (2016). MemGator — A portable concurrent memento aggregator: Cross-platform CLI and server binaries in Go," In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, June 2016, 243-244.
- Kelly, M., Alkwai, L. M., Alam, S., Nelson, M. L., Weigle, M. C., & Van de Sompel, H. (2017). Impact of URI Canonicalization on Memento Count, In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 303-304, arXiv:1703.03302.