

# Immutable Web Mobiles

## Web Archive Collection Zipped (WACZ)

Ed Summers

Investigative journalists, researchers, activists and everyday web practitioners are increasingly using web archives as part of their online information activities (Acker & Chalet, 2020). Web archives are used to revisit what web resources looked like in the past, but the trustworthiness of these representations is directly tied to the institutions that host the web archive itself. Questions about how a web archive was created, who created it, why they created it, and the completeness of the archive can be difficult, if not impossible, to ascertain. To make matters worse, the entanglement of web archives with the institutional infrastructures that publish them has worked against their portability and reuse in evidentiary contexts.

This talk will describe the design and implementation of the Web Archive Collection Zipped (WACZ) document format (Kreymer & Summers, 2021) which enables the packaging of heterogeneous and temporally fragile hypermedia networks as a discrete file for sharing and viewing when the original resource is no longer available. WACZ builds on the WARC format (International Organization for Standardization *The WARC format 1.1*, 2017) and the composability of the ZIP archive format in order to make web archive data and metadata available in a recognizable and reliable way.

Mechanisms for trust using public key cryptography, as well as random access using standard HTTP (Fielding & Reschke, 2014) publishing will be described and demonstrated using examples from the Webrecorder, Internet Archive and the Harvard Law Library (Zittrain, Bowers, & Stanton, 2021). The ability to create, share and use trustworthy web archives is critically important for them to circulate as immutable mobiles (Latour, 1999) on information networks.

### References

- Acker, A., & Chalet, M. (2020). The weaponization of web archives: Data craft and COVID-19 publics. *The Harvard-Kennedy School Misinformation Review*, 1. Retrieved from [https://misinforeview.hks.harvard.edu/wp-content/uploads/2020/09/acker\\_weaponization\\_web\\_archives\\_Covid19\\_20200927-1.pdf](https://misinforeview.hks.harvard.edu/wp-content/uploads/2020/09/acker_weaponization_web_archives_Covid19_20200927-1.pdf)
- Fielding, R., & Reschke, J. (2014). *Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content* (No. 7231). Internet Engineering Task Force. Retrieved

- from <https://datatracker.ietf.org/doc/html/rfc7231>
- International Organization for Standardization *The WARC format 1.1*. (2017). (No. ISO 28500:2017). Retrieved from <https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/>
- Kreymer, I., & Summers, E. (2021, June 3). Web Archive Collection Zipped (Version 1.1.1). Webrecorder. Retrieved from <https://specs.webrecorder.net/wacz/1.1.1/>
- Latour, B. (1999). Pandora's hope: Essays on the reality of science studies. In. Harvard University Press.
- Zittrain, J. L., Bowers, J., & Stanton, C. (2021). *The Paper of Record Meets an Ephemeral Web: An Examination of Linkrot and Content Drift within The New York Times* (SSRN Scholarly Paper No. ID 3833133). Rochester, NY: Social Science Research Network. <https://doi.org/10.2139/ssrn.3833133>