

Web Archives

Web archives are invaluable resources for hackers, journalists and other history enthusiasts. They provide a glimpse into the past, capturing snapshots of websites at various points in time. By leveraging these archives, we can uncover valuable insights, unravel historical data breaches, and gain a deeper understanding of an organization's online presence over the years.

Let's start with the most famous one.

Wayback machine - <https://web.archive.org/>

Historical Robots.txt file - https://web.archive.org/web/*/cnn.com/robots.txt

- **Archive.org Domain Registration Data** - web.archive.org

-> <https://web.archive.org/web/http://who.is/whois/domain.com>

-> <https://web.archive.org/web/https://whois.domaintools.com/domain.com>

-> <https://web.archive.org/web/https://whoxy.com/domain.com>

-> <https://web.archive.org/web/https://domainbigdata.com/domain.com>

-> <https://web.archive.org/web/https://whoisology.com/domain.com>

Next one we have is Archive Today

Archive Today - <https://archive.is/>

Memento web - <https://timetravel.mementoweb.org/list/19991212110000/http://domain.com>

Library of Congress - https://webarchive.loc.gov/all*/http://domain.com

Portugese Web Archive - <https://arquivo.pt/?l=en>

We can also check in google cache, because if a website has taken down something recently and google had not crawled the changed content, then it might holds a copy of the previous content in its cache, which could be beneficial for us.

Search Engine Cache - <https://web.cache.googleusercontent.com/search?q=cache:domain.com>
