

Remember, there is no single correct way to archive your own web content or the content of your chosen community. Your project goals and long-term information needs will determine what functionality your web archive needs to have. Depending on those goals and needs, and on your available resources (both financial and technical), you could do something as simple as a series of screenshots and webpage downloads from your browser in a set of nested folders, or as complex as a regularly scheduled site crawl with enhanced metadata.

You can't break the internet by archiving it "wrong"! With such a big cultural data loss at stake, doing something is always better than doing nothing.

You have the freedom to explore tools and determine what works best for you with your available resources. Just make sure that your chosen web archiving program is sustainable, consistent, and documented.

What Is a Web Archive?

- The Society of American Archivists' Dictionary of Archives Terminology definition: <https://dictionary.archivists.org/entry/web-archiving.html>
- Internet Archive's definition: <https://support.archive-it.org/hc/en-us/articles/360041674111-What-is-web-archiving>
- Wikipedia's list of worldwide archiving initiatives: https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives
 - If there is a web archiving initiative with the same focus as your organization, you may be able to join in

Why Is Web Archiving Important?

- Chapekis, A., Bestvater, S., Remy, E., & Rivero, G. (2024, May 17). *When online content disappears*. Pew Research Center. <https://www.pewresearch.org/data-labs/2024/05/17/when-online-content-disappears/>
- Freeland, C., Messarra, L., & Ziskina, J. (Eds.). (2024, October 30). *Vanishing culture: A report on our fragile cultural record*. Internet Archive. <https://blog.archive.org/2024/10/30/vanishing-culture-a-report-on-our-fragile-cultural-record/>
- Rosenthal, D.S.H. (2017, December 7). *Losing the battle to archive the web*. Digital Preservation Coalition. <https://www.dpconline.org/blog/wdpd/losing-the-battle-to-archive-the-web>

- Stokel-Walker, C. (2024, September 16). *We're losing our digital history. Can the internet archive save it?* BBC News. <https://www.bbc.com/future/article/20240912-the-archivists-battling-to-save-the-internet>

How Does Web Archiving Work?

- Virginia Tech University Libraries guide to web crawling: <https://guides.lib.vt.edu/webarchiving>
- Archive-It's workflow diagram: <https://support.archive-it.org/hc/en-us/articles/360041674111-What-is-web-archiving>

Where Do I Start?

- Determine your project scope
 - A presentation on identifying needs and goals for a web archiving initiative from the Metropolitan New York Library Council: <https://www.youtube.com/watch?v=4bS4WV8iQc0>
- Create policies for collecting and/or archiving web content
 - The Library of Congress's web archiving collections policy statement: <https://www.loc.gov/acq/devpol/webarchive.pdf>

How Do I Make Sure That My Website Is Archivable?

- Check your site through Archive Ready: <https://archiveready.com/>
- Library of Congress guidelines for creating preservable websites: <https://www.loc.gov/programs/web-archiving/for-site-owners/creating-preservable-websites/>

Specific Tools and Vendors

- Wayback Machine "Save Page Now" tool: <https://web.archive.org/save/>
 - You can also use a browser extension that lets you right-click and save that way
- Library of Congress web content request info and contact form link: <https://www.loc.gov/programs/web-archiving/about-this-program/frequently-asked-questions/>
- Community Webs application page: <https://communitywebs.archive-it.org/apply/>

- Archive-It paid products and services page: <https://archive-it.org/products-and-services/>
 - If you use programs such as Archive-It and Conifer that can store data for you, you should aim to do scheduled (2 times a year) downloads as a backup
- Archive-It sponsored program for “mission-aligned but financially under-resourced organizations”: <https://archive-it.org/blog/archive-it-sponsored/>
- Conifer web archiving service: <https://conifer.rhizome.org/>
 - A brief tutorial on using Conifer from the Norfolk Record Office: <https://www.youtube.com/watch?v=Y6eGu6LmkP4>
- ArchiveWeb.page web archiving tool: <https://webrecorder.net/archivewebpage/>
- ReplayWeb.page web archive viewer: <https://webrecorder.net/replaywebpage/>

Cataloging

- Library of Congress recommended formats statement for web archives, including basic metadata: <https://www.loc.gov/preservation/resources/rfs/webarchives.html>
- OCLC Research Library Partnership Web Archiving Metadata Working Group recommendations for descriptive metadata when cataloging web content: <https://doi.org/10.25333/C3005C>

Social Media Archiving

- Library of Congress and Twitter case study:
 - Raymond, M. (2010, April 14). *How tweet it is!: Library acquires entire Twitter Archive*. The Library of Congress. <https://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/>
 - Osterberg, G. (2017, December 26). *Update on the Twitter archive at the Library of Congress*. The Library of Congress. <https://blogs.loc.gov/loc/2017/12/update-on-the-twitter-archive-at-the-library-of-congress-2/>
- Request data download for your personal or organizational accounts:
 - Facebook: <https://www.facebook.com/help/1701730696756992>
 - Instagram: https://help.instagram.com/181231772500920?helpref=about_content
 - Twitter/X: <https://help.x.com/en/managing-your-account/accessing-your-x-data>

- YouTube:
<https://support.google.com/accounts/answer/3024190?sjid=13136367528540152585-NA>
- There are third-party paid services that export post comments, such as
<https://exportcomments.com/>

More Advanced Tools (these may require Python, JavaScript, Linux, etc.)

- Webrecorder's developer tools page: <https://webrecorder.net/developer-tools/>
- Documenting the Now's list of social media tools on GitHub:
<https://github.com/DocNow/awesome-social-media-archiving>
- Web Curator Tool, developed by National Library of New Zealand and National Library of the Netherlands: <https://webcuratortool.org/>
- NetArchive Suite, developed by the Royal Danish Library, National Library of France, and National Library of Austria: <https://github.com/netarchivesuite>
- Minet social media data gathering tools: <https://github.com/medialab/minet>

Further Reading and Viewing

- Digital Preservation Coalition's community Archives Toolkit Matrix with three levels of website preservation: <https://www.dpconline.org/digipres/implement-digipres/community-archives-dp-toolkit/community-archives-toolkit-matrix>
- The International Internet Preservation Consortium (IIPC)'s YouTube page includes videos of their 2024 conference sessions on a variety of web archiving topics:
<https://www.youtube.com/@iipc8855>
- OCLC Research Library Partnership Web Archiving Metadata Working Group review of harvesting tools, published in 2018: <https://doi.org/10.25333/C37H0T>