Web Archives @ Michigan State University FAQ

For Researchers

How do I know I am looking at an archived web page, and not the live web?

Web sites archived by the Michigan State University Archives & Historical Collections will always have a yellow banner at the top of the page with following:

'You are viewing an archived web page, collected at the request of Michigan State University using Archive-It. This page was captured on HH:MM:SS Month Day, Year, and is part of the Administration and Services Collection collection. The information on this web page may be out of date. See All versions of this archived page. '

Note: Some archived web sites contain internal links that are not captured by the Archive-It web crawler and will divert to the live web.

How do I remove my web page or photograph from the archive?

Please contact the University Archives staff regarding any issues about a captured web site or its parts. Also, if you would like for your web site or web domain not to be archived by the Michigan State University Archives & Historical Collections web crawler in the future, please contact the [University Archives staff](#).

How do I get my web site included in the MSU web archives?

If your web site is part of the MSU campus community and your web content is not currently being archived by the Internet Archive or the Archive-It Program, you can contact the [University Archives staff](#) to inquire about having your web site included in the UAHC web archive crawls.

How does the University Archives decide which web sites to preserve in the web archives?

The University Archives staff appraises web site for enduring historical value. A Web Archives Collection Plan will be posted soon with more information.

Does Archive-IT capture the date and time when a web page is updated?

No. The Archive-It web crawler can only take a snapshot of a web site either before or after a web site has changed. Currently, there is no way to exactly determine when and how a particular web site modifies its content. The Archive-It page banner displays when the page was crawled.

What's a robots.txt file and what's it have to do with your web archiving?

The robots.txt exclusion standard, is a way for a webmaster to direct a web crawler (aka robot or spider) not to crawl all or specified parts of their website. The webmaster places their request in

the form of a specific text file that is easily found on their website. Archive-It (like Google and most search engines) uses a robot to crawl and archive web pages. By default, the Archive-It crawler honors and respects all robots.txt exclusion requests. The UAHC may request that you add archive.org_bot as an allowed crawler so that we can capture all the files needed to correctly display the pages.

If I notice a page is displaying oddly or missing content in the web archives, what should I do?

Please let us know! There are too many pages for our staff to validate so any help in pointing out problems is greatly appreciated. You can contact the [University Archives staff](#).

For Website Developers

Can you provide guidance on how to create websites that are captured the best?

Check out the excellent [reference](#) created by Stanford University.