

8-13-2021

## The Internet Archive has been Fighting for 25 Years to Keep What's on the Web From Disappearing - and You Can Help

Kayla Harris

University of Dayton, kharris2@udayton.edu

Stephanie Shreffler

University of Dayton, sshreffler1@udayton.edu

Christina A. Beis

University of Dayton, cbeis1@udayton.edu

Follow this and additional works at: [https://ecommons.udayton.edu/imri\\_faculty\\_publications](https://ecommons.udayton.edu/imri_faculty_publications)



Part of the [Archival Science Commons](#), and the [Digital Humanities Commons](#)

---

### eCommons Citation

Kayla Harris (0000-0002-1672-3022), Stephanie Shreffler (0000-0002-2147-7495), and Christina A. Beis (0000-0003-4749-1292) (2021). The Internet Archive has been Fighting for 25 Years to Keep What's on the Web From Disappearing - and You Can Help. *The Conversation*. , [https://ecommons.udayton.edu/imri\\_faculty\\_publications/62](https://ecommons.udayton.edu/imri_faculty_publications/62)

This News Article is brought to you for free and open access by the The Marian Library at eCommons. It has been accepted for inclusion in Marian Library Faculty Publications by an authorized administrator of eCommons. For more information, please contact [mschlangen1@udayton.edu](mailto:mschlangen1@udayton.edu), [ecommons@udayton.edu](mailto:ecommons@udayton.edu).

Aug. 13, 2021

# The Internet Archive has been fighting for 25 years to keep what's on the web from disappearing – and you can help

---

**Kayla Harris**, Marian Library Librarian/Archivist; Associate Professor, University of Dayton

**Christina Beis**, Director of Collections Strategies & Services, Associate Professor, University Libraries, University of Dayton

**Stephanie Shreffler**, Collections Librarian/Archivist and Associate Professor, University Libraries, University of Dayton

---

This year the Internet Archive turns 25. It's best known for its pioneering role in archiving the internet through the Wayback Machine, which allows users to see how websites looked in the past.

Increasingly, much of daily life is conducted online. School, work, communication with friends and family, as well as news and images, are accessed through a variety of websites. Information that once was printed, physically mailed or kept in photo albums and notebooks may now be available only online. The COVID-19 pandemic has pushed even more interactions to the web.

You may not realize portions of the internet are constantly disappearing. As librarians and archivists, we strengthen collective memory by preserving materials that document the cultural heritage of society, including on the web. You can help us save the internet, too, as a citizen archivist.

## **Disappearing act**

People and organizations remove content from the web for a variety of reasons. Sometimes it's a result of changing internet culture, such as the recent shutdown of Yahoo Answers.

It can also be a result of following best practices for website design. When a website is updated, for example, the previous version is overwritten – unless it was archived.

Web archiving is the process of collecting, preserving and providing continued access to information on the internet. Often this work is done by librarians and archivists, with assistance from automated technology like web crawlers.

Web crawlers are programs that index web pages to make them available through search engines, or for long-term preservation. The Internet Archive, a nonprofit organization, uses thousands of computer servers to save multiple digital copies of these pages, requiring over 70 petabytes of data. It is funded through donations, grants and payments for its digitization services. Over 750 million web pages are captured per day in the Internet Archive’s Wayback Machine.

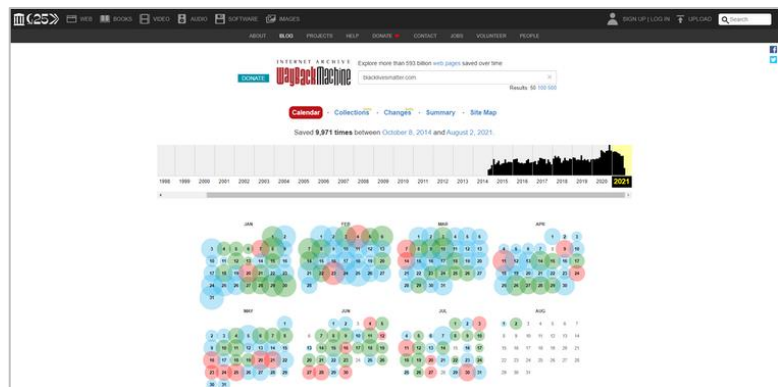
### Why archive?

In 2018, President Donald Trump wrongly claimed via Twitter that Google had promoted on its homepage President Barack Obama’s State of the Union address, but not his own. Archived versions of the Google homepage proved that Google had, in fact, highlighted Trump’s State of the Union address in the same manner. Multiple news outlets use the Internet Archive’s Wayback Machine as the source for fact-checking these types of claims, since screenshots alone can be easily altered.

A 2019 report from the Tow Center for Digital

Journalism examined the digital archiving practices and policies of newspapers, magazines and other news producers. The interviews revealed that many news media staff either do not have the resources to devote to archiving their work or misunderstand digital archiving by equating it to having a backup version.

When a news story disappeared from the Gawker website a year after the publication shut down, the Freedom of the Press Foundation became concerned with what might happen when wealthy individuals purchase websites with the intent to delete or censor the archives. It partnered with the Internet Archive to launch a web archive collection focused on preserving the



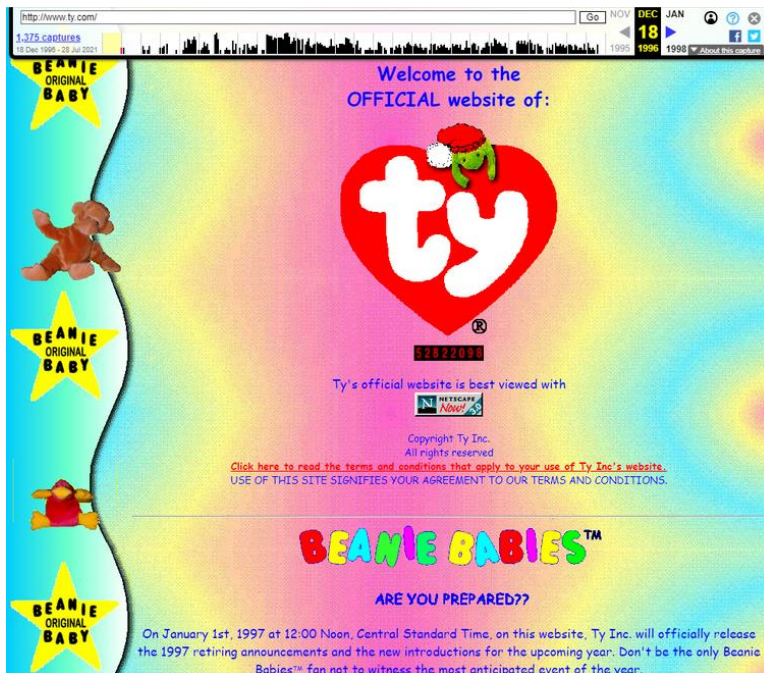
*The web crawls for blacklivesmatter.com in the Internet Archive’s Wayback Machine. Internet Archive Wayback Machine*

web archives of vulnerable news outlets – and to dissuade billionaires from purchasing such material to censor.

Archiving websites that document social justice issues, such as Black Lives Matter, helps explain these movements to people of the present and the future.

Archiving government websites promotes transparency and accountability. Especially during times of transition, government websites are vulnerable to deletion with changing political parties.

In 2017 the Library of Congress announced it would no longer archive every single tweet, because of Twitter’s growth as a communication tool. Twitter supplies the Library of Congress with the texts of tweets, not shared images or videos. Instead of comprehensive collecting, the Library of Congress now archives only tweets of significant national importance.



Screen capture from the Dec. 18, 1996, archived version of the Ty website, creator of Beanie Babies, in the Internet Archive’s Wayback Machine. Internet Archive Wayback Machine

Archived websites that document the culture and history of the internet, like the Geocities Gallery, not only are fun to look at but illustrate the ways early websites were created and used by individuals.

### Citizen archivists

Archiving the internet is a monumental task, one that librarians and archivists cannot do alone. Anyone can be a citizen archivist and preserve history through the Internet Archive’s Wayback Machine.

The “Save Page Now” feature allows anyone to freely archive a single, public website page. Bear in mind, some websites prevent web crawling and archiving through special coding or by requiring a login to the site. This may be due to sensitive content or the personal preference of the web developer.

Local cultural heritage institutions, such as libraries, archives and museums, are also actively archiving the internet. Over 800 institutions use Archive-It, a tool from the Internet Archive, to create archived web collections. At the University of Dayton we curate collections related to our Catholic and Marianist heritage, from Catholic blogs to stories of the Virgin Mary in the news.

Through its Spontaneous Event collections, Archive-It partners with organizations and individuals to create collections of “web content related to a specific event, capturing at risk content during times of crisis.”

Similarly, it created the Community Webs program, in partnership with the Institute of Museum and Library Services, to help public libraries create collections of archived web content relevant to local communities.

The websites of today are the historical evidence of tomorrow, but only if they are archived. If they are lost, we will lose crucial information about corporate and government decisions, modern communication methods such as social media, and social movements with significant online presences, such as Black Lives Matter and #MeToo.

Together with librarians and archivists, you can help ensure the survival of this evidence and save internet history.