

# Is the internet really forever? Link rot requires infrastructure changes and human collaboration to solve

May 29 2024, by Cesareo Contreras

---



Credit: Pixabay/CC0 Public Domain

It's happened to all of us at one point while browsing the web—somewhere along the way, you click on a broken link and get a message saying that the web page you are looking for doesn't exist.

It's as common as it is frustrating, and it only seems like the problem will get worse as the internet continues to expand and old web pages get migrated or abandoned.

There's even a name for the issue. It's called link rot, a term that dates to the '90s when the internet rose to prominence.

Earlier this month, Pew Research Center [released a report digging deep on the issue](#), finding that a third of web pages that existed in 2013 can no longer be accessed.

Here are some other insights Pew uncovered:

- "23% of news web pages contain at least one broken link, as do 21% of webpages from government sites."
- "54% of Wikipedia pages contain at least one link in their 'References' section that points to a page that no longer exists."
- "Nearly one-in-five tweets are no longer publicly visible on the site just months after being posted."

Joseph Reagle, associate professor of communication studies at Northeastern University, says the problem starts with the infrastructure of URL technologies, which stand for uniform resource locator.

URLs serve as address points for web pages on the internet similar to addresses for physical places like your home or work. URLs are great because they allow people to easily find websites, but the issue is they can be easily broken, he says.

In the '90s, Reagle worked with Tim Berners-Lee, largely credited with inventing the internet, at the World Web Consortium as a policy analyst. The issues around URLs were spoken of at length.

"We knew, for instance, that URLs are not maintained very well. If you're an organization or a company and you decide to reorganize or decide you are going to change platforms, all the URLs typically break."

In the web's early days, internet technologists examined the idea of using alternatives for the URL system. One proposal was instead to use URN-based technologies, which stand for uniform resource name, that would work similarly to the ISBN system used to catalog books, Reagle says.

But the problem there is that some larger organization would be in charge of managing it. The ISBN system is managed by the International ISBN agency, an entity that was appointed by the International Organization for Standardization.

"So you are stuck with two problems," Reagle says. "Either you let everyone create their URLs and manage their resources, and they tend to be really bad about it over time, or you create centralized repositories with permanent identities, but setting those up is costly and difficult to maintain."

The URL system has thus become the primary way people interface with the web, he notes, and issues around link rot remain.

"People do raise the issue from time to time. It gets a little attention, and then the world moves on," Reagle says. "There's been attempts at solutions, but the problems still persist."

Archivist organizations have come out of the woodwork to help solve these issues. A few notable projects include the Wayback Machine, archive.today and perma.cc, which allow people to access old versions of web pages that are no longer active and archive new [web pages](#) themselves.

But these services largely exist precariously and in the shadows, Regale notes, largely by small groups of people with a deep interest in online preservation.

These efforts also require individual users to help build out their databases, which can be seen as a big ask and is insufficient in adequately archiving large swaths of the web.

"They are all a little bit different, and they all are not perfect," he says. "Perma.cc and other similar programs require people to proactively, say, "Hey, make a copy of this page." and not everyone's going to do that. There are vast amounts of the web that are not on Perma.cc."

These issues go beyond infrastructure and human collaboration challenges. There are also issues around copyright concerns and what legal protections individuals have in working to preserve the web, Regale adds.

That's where the federal government could help play a role.

"I could imagine [Congress] passing a law that for instance gave people safe harbor provisions to people that were archiving content for purposes of education or research," he says.

*This story is republished courtesy of Northeastern Global News [news.northeastern.edu](https://news.northeastern.edu).*

Provided by Northeastern University

Citation: Is the internet really forever? Link rot requires infrastructure changes and human collaboration to solve (2024, May 29) retrieved 10 August 2024 from <https://techxplore.com/news/2024-05-internet-link-requires-infrastructure-human.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.