

38% of content on the Internet has 'disappeared' in the past 10 years

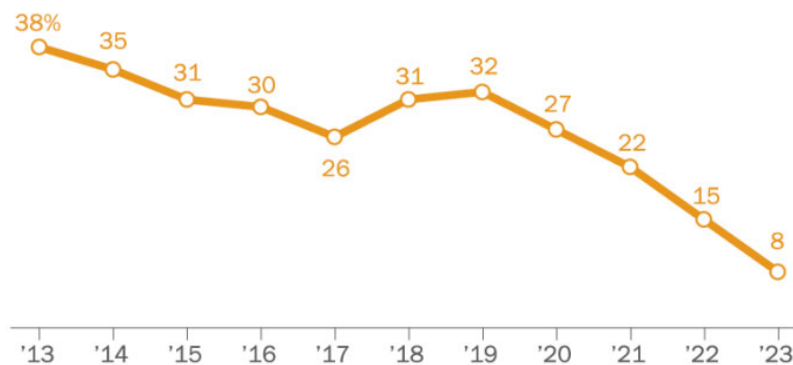
According to new research recently shared by Pew Research Center, up to 38% of content posted on the Internet on websites since 2013 no longer exists.

At the end of the study, in 2023, 8% of the content created had disappeared in just one year. This phenomenon was named "digital decay" by the authors.

Specifically, 23% of news sites have at least one broken link. The data was released by researchers after surveying more than 500 thousand pages of more than 2 thousand websites classified as news sites.

38% of webpages from 2013 are no longer accessible

% of links from each year that are no longer accessible as of October 2023



Source: Pew Research Center analysis of a random selection of URLs collected by the Common Crawl web repository (n=999,989) and checked using page and DNS response codes. Web pages defined as inaccessible if they returned a status code of 204, 400, 404, 410, 500, 501, 502, 503, 523 or did not return a valid status code.

"When Online Content Disappears"

PEW RESEARCH CENTER

For government websites, this number is 21%. Among them, small local pages, such as district or city level, often have broken links the most. Most happen on pages with certain security modes such as https://, 6% of which are related to a certain pdf file, 16% are pointed to another URL that is not consistent with the original recorded content. head.

With free information sites like Wikipedia, the situation of broken links is much higher. 54% of the 50 thousand pages of information are broken in the References section.

The situation of digital decay is even stronger on social networks. For example, Twitter or now X nearly 1/5 of tweets are no longer set to "public" free viewing mode after just a few months. 60% of them may be due to the account owner being locked, deleted or converted to private mode. The remaining 40% are account owners who delete their posts themselves.

More specifically, 1% of tweets are deleted immediately after 1 hour of posting, 3% are deleted after 1 day, 10% are deleted after 1 week and 15% are deleted after 1 month.

Below are the most commonly seen group errors

1. 204 No Content
2. 400 Bad Requests
3. 404 Not Found
4. 410 Gone
5. 500 Internal Server Error
6. 501 Not Implemented
7. 502 Bad Gateway
8. 503 Service Unavailable
9. 523 Origin Is Unreachable

You finished reading the article "**38% of content on the Internet has 'disappeared' in the past 10 years**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.