INTERNET
ARCHIVE

USING THE WAYBACK MACHINE

This introduction video provides an overview for how to use the Wayback Machine, including information about searching by URL or keyword, understanding provenance, and saving your own pages, along with other features.

▶ 0:00 / 21:06   🔊   ⛶   ⋮
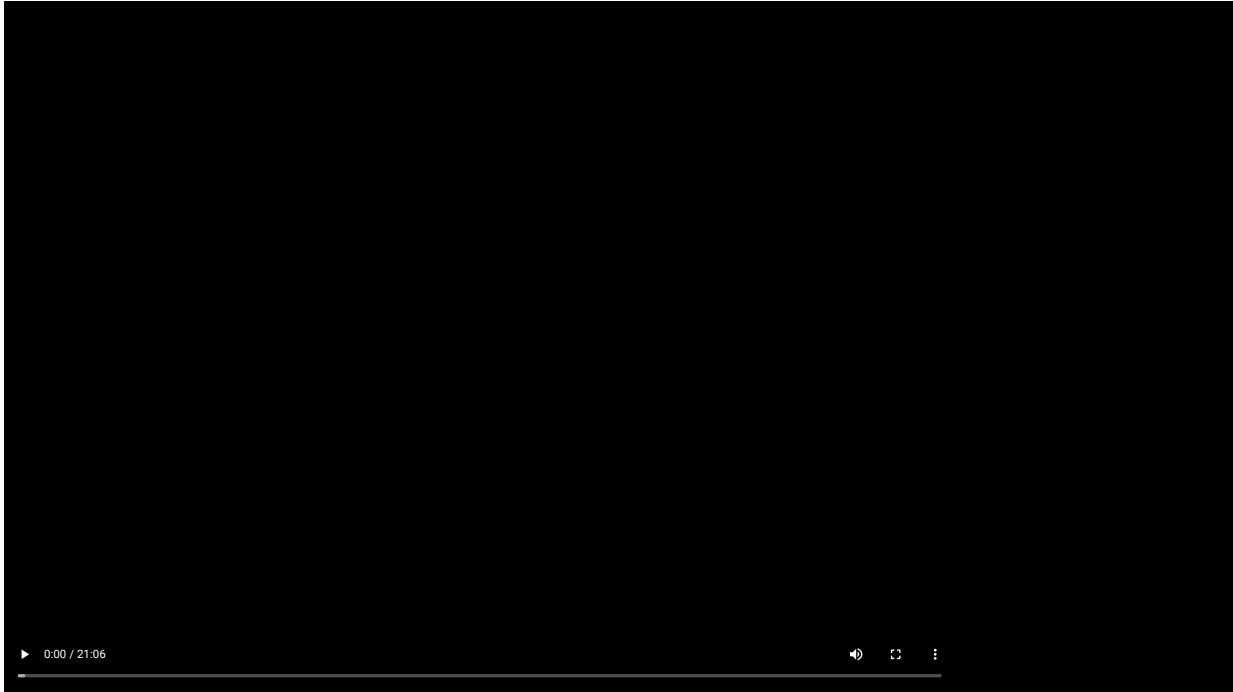
Can I link to old pages on the Wayback Machine?

Yes! The Wayback Machine is built so that it can be used and referenced. If you find an archived page that you would like to reference on your Web page or in an article, you can copy the URL. You can even use fuzzy URL matching and date specification... but that's a bit more advanced.

How can I use the Wayback Machine's Site Search to find websites?

The Site Search feature of the Wayback Machine is based on an index built by evaluating terms from hundreds of billions of links to the homepages of more than 350 million sites. Search results are ranked by the number of captures in the Wayback and the number of relevant links to the site's homepage.

Can I search the Archive?

Using the Internet Archive Wayback Machine, it is possible to search for the names of sites contained in the Archive (URLs) and to specify date ranges for your search. We hope to implement a full text search engine at some point in the future.

Why isn't the site I'm looking for in the archive?

Some sites may not be included because the automated crawlers were unaware of their existence at the time of the crawl. It's also possible that some sites were not archived because they were password protected, blocked by robots.txt, or otherwise inaccessible to our automated systems. Site owners might have also requested that their sites be excluded from the Wayback Machine.

How can I exclude or remove my site's pages from the Wayback Machine?

If you would like to submit a request for archives of your site or account to be excluded from web.archive.org, send us a request to info@archive.org and indicate:

- the URL or URLs of the material
- the time period that you wish to have excluded
- the time period during which you had control of the site or relevant user account (if applicable) and
- any other information that you think would be helpful for us to better understand your request.

This will initiate a review by our team. We do not make any guarantees beforehand about the outcome of a request.

How can I use the Wayback Machine's Site Search to find websites?

The Site Search feature of the Wayback Machine is based on an index built by evaluating terms from hundreds of billions of links to the homepages of more than 350 million sites. Search results are ranked by the number of captures in the Wayback and the number of relevant links to the site's homepage.

How can I get a copy of the pages on my Web site? If my site got hacked or damaged, could I get a backup from the Archive?'

Our terms of use do not cover backups for the general public. However, you may use the Internet Archive Wayback Machine to locate and access archived versions of a site to which you own the rights. We can't guarantee that your site has been or will be archived. We can no longer offer the service to pack up sites that have been lost.

Can I add pages to the Wayback Machine?

On https://archive.org/web you can use the "Save Page Now" feature to save a specific page one time. This does not currently add the URL to any future crawls nor does it save more than that one page. It does not save multiple pages, directories or entire sites.

Where is the rest of the archived site? Why am I getting broken or gray images on a site?

Broken images occur when the images are not available on our servers. Usually this means that we did not archive them.

You can tell if the image or link you are looking for is in the Wayback Machine by entering the image or link's URL into the Wayback Machine search box. Whatever archives we have are viewable in the Wayback Machine.

The best way to see all the files we have archived of the site is: http://web.archive.org/*/www.yoursite.com/*

There is a 3-10 hour lag time between the time a site is crawled and when it appears in the Wayback Machine.

Why are some sites harder to archive than others?

If you look at our collection of archived sites, you will find some broken pages, missing graphics, and some sites that aren't archived at all. Some of the things that may cause this are:

Robots.txt — A site's robots.txt document may have prevented the crawling of a site.
Javascript — Javascript elements are often hard to archive, but especially if they generate links without having the full name in the page. Plus, if javascript needs to contact the originating server in order to work, it will fail when archived.
Server side image maps — Like any functionality on the web, if it needs to contact the originating server in order to work, it will fail when archived.
Orphan pages — If there are no links to your pages, the robot won't find it (the robots don't enter queries in search boxes.)
As a general rule of thumb, simple html is the easiest to archive.
Can I find sites by searching for words that are in their pages?

No, at least not yet. Site Search for the Wayback Machine will help you find the homepages of sites, based on words people have used to describe those sites, as opposed to words that appear on pages from sites.

Can I still find sites in the Wayback Machine if I just know the URL?

Yes, just enter a domain or URL the way you have in the past and press the "Browse History" button.

Why are some of the dots on the calendar page different colors?

We color the dots, and links, associated with individual web captures, or multiple web captures, for a given day. Blue means the web server result code the crawler got for the related capture was a 2nn (good); Green means the crawlers got a status code 3nn (redirect); Orange means the crawler got a status code 4nn (client error), and Red means the crawler saw a 5nn (server error). Most of the time you will probably want to select the blue dots or links.

How does the Wayback Machine behave with Javascript turned off?

If you have Javascript turned off, images and links will be from the live web, not from our archive of old Web files.

How did I end up on the live version of a site? or I clicked on X date, but now I am on Y date, how is that possible?

Not every date for every site archived is 100% complete. When you are surfing an incomplete archived site the Wayback Machine will grab the closest available date to the one you are in for the links that are missing. In the event that we do not have the link archived at all, the Wayback Machine will look for the link on the live web and grab it if available. Pay attention to the date code embedded in the archived url. This is the list of numbers in the middle; it translates as yyyymmddhhmmss. For example in this url http://web.archive.org/web/20000229123340/http://www.yahoo.com/ the date the site was crawled was Feb 29, 2000 at 12:33 and 40 seconds.

You can see a listing of the dates of the specific URL by replacing the date code with an asterisk (*), ie: http://web.archive.org/*/www.yoursite.com

How do I cite Wayback Machine urls in MLA format?

This question is a newer one. We asked MLA to help us with how to cite an archived URL in correct format. They did say that there is no established format for resources like the Wayback Machine, but it's best to err on the side of more information. You should cite the webpage as you would normally, and then give the Wayback Machine information. They provided the following example: McDonald, R. C. "Basic Canary Care." _Robirda Online_. 12 Sept. 2004. 18 Dec. 2006 [http://www.robirda.com/cancare.html]. _Internet Archive_. [http://web.archive.org/web/20041009202820/http://www.robirda.com/cancare.html]. They added that if the date that the information was updated is missing, one can use the closest date in the Wayback Machine. Then comes the date when the page is retrieved and the original URL. Neither URL should be underlined in the bibliography itself. Thanks MLA!

How can I get pages authenticated from the Wayback Machine? How can I use the pages in court? While the Wayback Machine tool was not expressly designed with legal use in mind, we receive regular requests for certified records for use in legal proceedings. Our affidavit request procedure can be found here. Please review that information including our standard affidavit and the legal request FAQ section linked there to prior to contacting us.

Some sites are not available because of robots.txt or other exclusions. What does that mean?

Such sites may have been excluded from the Wayback Machine due to a robots.txt file on the site or at a site owner's direct request.

How can I get my site included in the Wayback Machine?

Much of our archived web data comes from our own crawls or from Alexa Internet's crawls. Neither organization has a "crawl my site now!" submission process. Internet Archive's crawls tend to find sites that are well linked from other sites. The best way to ensure that we find your web site is to make sure it is included in online directories and that similar/related sites link to you.

Alexa Internet uses its own methods to discover sites to crawl. It may be helpful to install the free Alexa toolbar and visit the site you want crawled to make sure they know about it.

Regardless of who is crawling the site, you should ensure that your site's 'robots.txt' rules and in-page META robots directives do not tell crawlers to avoid your site.

What is the Archive-It service of the Internet Archive Wayback Machine?

For information on the Archive-It subscription service that allows institutions to build and preserve collections of born digital content, see https://www.archive.org/about/faqs.php#Archive-It

Was this helpful?   👍 Yes   No 👎   266 / 113

Search

[   ] Search

Accounts
Archive-it Service
Archive.org
Archive.org General Information
Audio and Music Items Information
Books and Texts Information
Borrow a Book from Archive.org
Charges & Refunds

Collections
Company Matching & Amazon Smile
Disability Access
Donating Physical Items to the Archive
Donating to the Internet Archive
Donations
Downloading
Favorites
Files, Formats, and Derivatives
Forums
General Interest
Kid-Friendly
Managing and Editing Your Items
Media Players
Most Frequently Asked Questions
Movies and Videos
Organization Information
Reporting problems or errors
Search
Software
Tax Information
Technical Information
The Wayback Machine
Uploading
Wayback Machine & Web Archiving

Recent Posts
Donate Books App for iOS and Android
How to archive your Tweets with the Wayback Machine
Media Types for Donations
Does the Internet Archive Have My Media
File Formats