



CDMM Tutorial

Internet Archive's "Save Page Now"

Tutorial by
Asger Harlung
CDMM, 2022

CDMM

Centre for Digital Methods and Media
<https://cc.au.dk/cdmm/>

Save a Web Page to Internet Archive

This tutorial will show you step by step how to save a web page to the The Internet Archive.

Minor limitations will be mentioned at the end of the tutorial.

The tutorial can also be downloaded as PDF.

Save a Web Page to Internet Archive

An archived web page is less likely to disappear than a web page found on the "live" web.

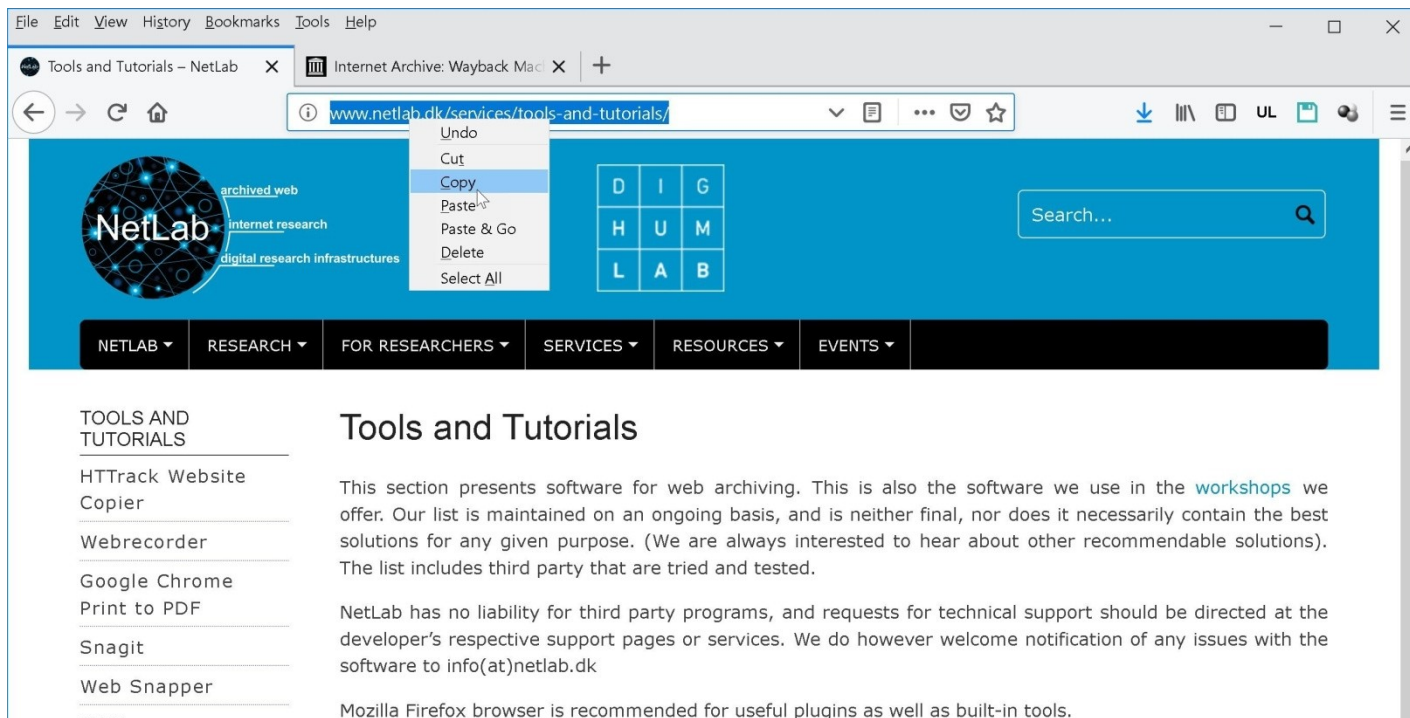
Internet Archive's "Save page now" is useful for preservation, but also for references that are more stable than the original URLs of the pages you would like to reference.

Advantages

- Easy way to preserve content immediately.
- More stable as a reference than original URL.
- Works with most web pages.

Here is how it works:

1) Copy the URL for the page to preserve



The screenshot shows a web browser window with the following elements:

- Browser Tabs:** 'Tools and Tutorials - NetLab' and 'Internet Archive: Wayback Mac'.
- Address Bar:** 'www.netlab.dk/services/tools-and-tutorials/'
- Context Menu:** Open over the address bar with 'Copy' selected.
- Page Header:** NetLab logo (archived web, internet research, digital research infrastructures), a search bar, and a navigation menu (NETLAB, RESEARCH, FOR RESEARCHERS, SERVICES, RESOURCES, EVENTS).
- Main Content:**
 - Section Header:** 'Tools and Tutorials'
 - Text:** 'This section presents software for web archiving. This is also the software we use in the [workshops](#) we offer. Our list is maintained on an ongoing basis, and is neither final, nor does it necessarily contain the best solutions for any given purpose. (We are always interested to hear about other recommendable solutions). The list includes third party that are tried and tested.'
 - Text:** 'NetLab has no liability for third party programs, and requests for technical support should be directed at the developer's respective support pages or services. We do however welcome notification of any issues with the software to [info\(at\)netlab.dk](mailto:info(at)netlab.dk)'
 - Text:** 'Mozilla Firefox browser is recommended for useful plugins as well as built-in tools.'
- Left Sidebar:** A list of tools and tutorials: 'TOOLS AND TUTORIALS', 'HTTrack Website Copier', 'Webrecorder', 'Google Chrome', 'Print to PDF', 'Snagit', 'Web Snapper', and '.....'.

2) Go to archive.org/web

File Edit View History Bookmarks Tools Help

Tools and Tutorials - NetLab x Internet Archive: Wayback Ma x +

https://archive.org/web/

ABOUT CONTACT BLOG PROJECTS HELP DONATE JOBS VOLUNTEER PEOPLE

INTERNET ARCHIVE Explore more than 347 billion web pages saved over time

waybackmachine http:// BROWSE HISTORY

Find the Wayback Machine useful? DONATE

Tools

Wayback Machine Availability API
Build your own tools.

WordPress Broken Link Checker
Banish broken links from your blog.

404 Handler for Webmasters
Help users get where they were going.

Subscription Service

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. Visit [Archive-It to build and browse the collections.](#)

Save Page Now

http:// SAVE PAGE

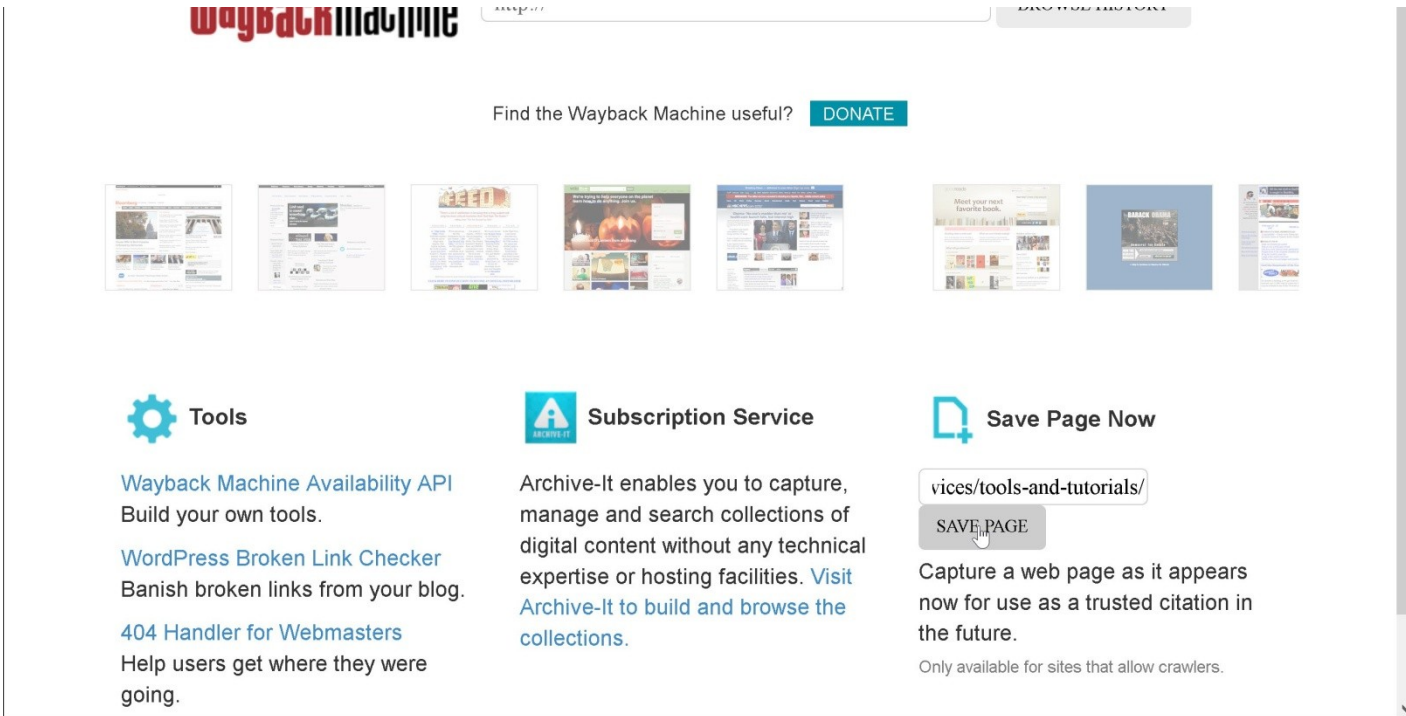
Capture a web page as it appears now for use as a trusted citation in the future.

Only available for sites that allow crawlers.

**“Save page now”
is in the bottom
right corner**


**... notice the
small print; this
may not work on
all websites.**

3) Paste URL and click "Save Page"



The screenshot shows the Wayback Machine homepage. At the top, the Wayback Machine logo is on the left, and a search bar is on the right. Below the logo, there is a question "Find the Wayback Machine useful?" followed by a teal "DONATE" button. A row of ten small thumbnail images of various web pages is displayed below. At the bottom, there are three main sections: "Tools" with a gear icon, "Subscription Service" with an Archive-It logo, and "Save Page Now" with a blue square icon. The "Save Page Now" section has a text input field containing "vices/tools-and-tutorials/" and a "SAVE PAGE" button with a mouse cursor over it. Below the button, there is a description of the service and a note that it is only available for sites that allow crawlers.


Find the Wayback Machine useful? [DONATE](#)

 **Tools**


[Wayback Machine Availability API](#)
Build your own tools.

[WordPress Broken Link Checker](#)
Banish broken links from your blog.

[404 Handler for Webmasters](#)
Help users get where they were going.

 **Subscription Service**

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. [Visit Archive-It to build and browse the collections.](#)

 **Save Page Now**

[SAVE PAGE](#)

Capture a web page as it appears now for use as a trusted citation in the future.

Only available for sites that allow crawlers.

4) The web page is now saving...

The screenshot shows a web browser window with the address bar displaying `https://web.archive.org/save/http://www.netlab.dk/services...`. A white dialog box is centered on the screen with the following text:

Saving page now...

<http://www.netlab.dk/services/tools-and-tutorials/>

As it appears live February 13, 2019 1:45:20 PM UTC

.....

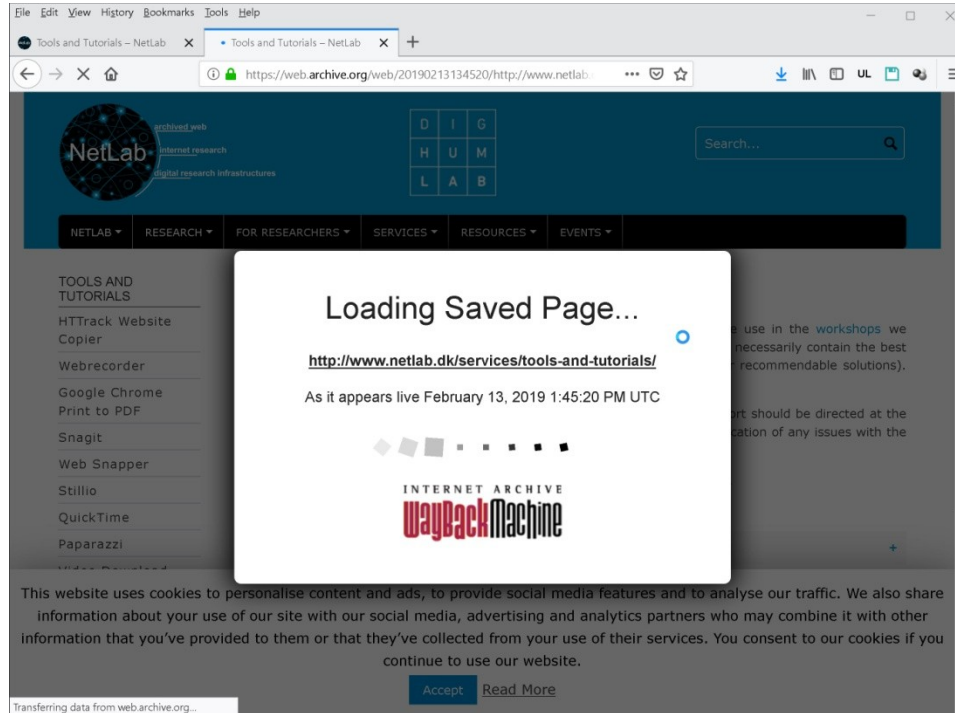
INTERNET ARCHIVE
WayBackMachine

The background page is titled "Tools and Tutorials" and features a sidebar with the following items:

- TOOLS AND TUTORIALS
- HTTrack Website Copier
- Webrecorder
- Google Chrome
- Print to PDF
- Snagit
- Web Snapper
- Stillo
- QuickTime
- Paparazzi
- Video Download-Helper

At the bottom of the page, there is a cookie consent notice with "Accept" and "Read More" buttons. A status bar at the very bottom indicates "Transferring data from web.archive.org..."

5) After saving the copied page will load:



6) Archived copy of the page:

The screenshot shows a web browser window with the address bar displaying the URL: <https://web.archive.org/web/20190213134520/http://www.netlab.dk/services/tools-and-tutorials/>. The page is an archived version of the NetLab website, as indicated by the Wayback Machine logo and the "14 captures" information.

The main content area features a blue header with the NetLab logo and the text "archived web", "internet research", and "digital research infrastructures". A search bar is visible on the right side of the header. Below the header is a navigation menu with the following items: NETLAB, RESEARCH, FOR RESEARCHERS, SERVICES, RESOURCES, and EVENTS.

The main heading is "Tools and Tutorials". The text below the heading reads: "This section presents software for web archiving. This is also the software we use in the [workshops](#) we offer. Our list is maintained on an ongoing basis, and is neither final, nor does it necessarily contain the best solutions for any given purpose. (We are always interested to hear about other recommendable solutions). The list includes third party that are tried and tested."

Below the text, there is a list of tools and tutorials:

- HTTrack Website Copier
- Webrecorder
- Google Chrome
- Print to PDF

The page also includes a disclaimer: "NetLab has no liability for third party programs, and requests for technical support should be directed at the developer's respective support pages or services. We do however welcome notification of any issues with the..."

7) Expanded view of full archive URL

The screenshot shows a web browser window with two tabs. The active tab is titled "Tools and Tutorials - NetLab". The address bar contains the URL: <https://web.archive.org/web/20190213134520/http://www.netlab.dk/services/tools-and-tutorials/>. Below the address bar, the Wayback Machine interface is visible, showing the original URL <http://www.netlab.dk/services/tools-and-tutorials/> and a "GO" button. To the right, there is a calendar navigation showing the current date as February 13, 2019, with options for 2018, 2019, and 2020. Below the calendar, there is a "14 captures" section with a date range from 30 Jun 2017 to 13 Feb 2019. The main content area has a blue header with the NetLab logo and the text "archived web", "internet research", and "digital research infrastructures". To the right of the logo is a grid of letters: D I G, H U M, L A B. Below the header is a navigation menu with items: NETLAB, RESEARCH, FOR RESEARCHERS, SERVICES, RESOURCES, and EVENTS.

TOOLS AND TUTORIALS

[HTTrack Website Copier](#)

[Webrecorder](#)

[Google Chrome](#)

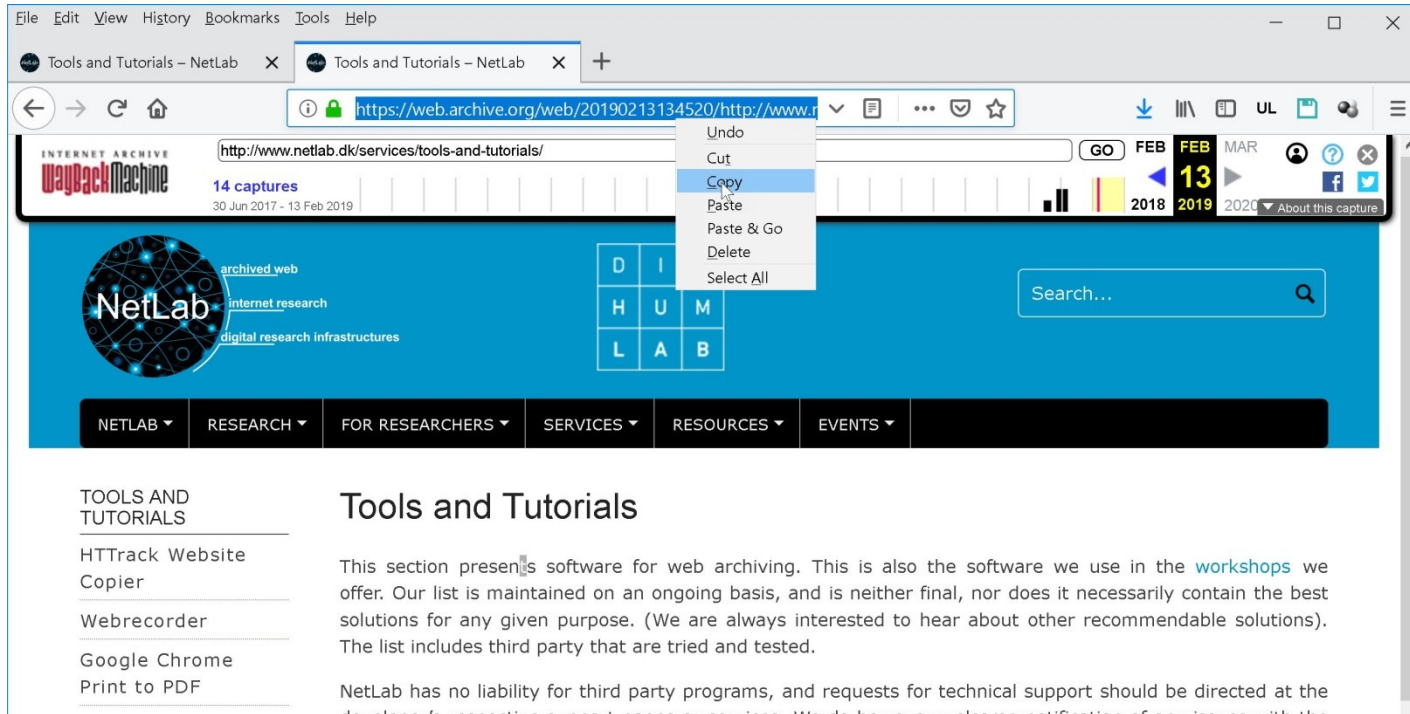
[Print to PDF](#)

Tools and Tutorials

This section presents software for web archiving. This is also the software we offer. Our list is maintained on an ongoing basis, and is neither final, nor definitive solutions for any given purpose. (We are always interested to hear about new solutions. The list includes third party that are tried and tested.

NetLab has no liability for third party programs, and requests for technical support should be directed to the respective support pages or services. We do however welcome

8) Copying the archive URL



The screenshot shows a web browser window with the following elements:

- Browser Tabs:** Two tabs titled "Tools and Tutorials - NetLab".
- Address Bar:** The URL is `https://web.archive.org/web/2019021314520/http://www.netlab.dk/services/tools-and-tutorials/`. A context menu is open over the address bar with the following options: Undo, Cut, Copy (highlighted), Paste, Paste & Go, Delete, and Select All.
- Wayback Machine Header:** Shows "INTERNET ARCHIVE WayBackMachine" and "14 captures" from "30 Jun 2017 - 13 Feb 2019".
- NetLab Logo:** A circular logo with the text "NetLab" and "archived web internet research digital research infrastructures".
- Navigation Menu:** A dark bar with dropdown menus for "NETLAB", "RESEARCH", "FOR RESEARCHERS", "SERVICES", "RESOURCES", and "EVENTS".
- Page Content:**
 - TOOLS AND TUTORIALS** (Section Header)
 - HTTrack Website Copier**
 - Webrecorder**
 - Google Chrome Print to PDF**
 - Tools and Tutorials** (Section Header)
 - This section presents software for web archiving. This is also the software we use in the [workshops](#) we offer. Our list is maintained on an ongoing basis, and is neither final, nor does it necessarily contain the best solutions for any given purpose. (We are always interested to hear about other recommendable solutions). The list includes third party that are tried and tested.
 - NetLab has no liability for third party programs, and requests for technical support should be directed at the developer's respective support pages or services. We do however welcome notification of any issues with the

Archive URL Structure

The archived URL contains useful information for referencing purposes:

<https://web.archive.org/web/20190213134520/http://www.netlab.dk/services/tools-and-tutorials/>

<https://web.archive.org/web> means the copy is in
The Internet Archive.

The second part of the archived URL:

<https://web.archive.org/web/20190213134520/http://www.netlab.dk/services/tools-and-tutorials/>

[20190213134520/](https://web.archive.org/web/20190213134520/http://www.netlab.dk/services/tools-and-tutorials/) is a time stamp with the formula
YYYYMMDDhhmmss. Time of day (last six digits)
is GMT.

And the final part of the archived URL:

<https://web.archive.org/web/20190213134520/http://www.netlab.dk/services/tools-and-tutorials/>

<http://www.netlab.dk/services/tools-and-tutorials/> is the original URL for the web page that was copied.

Limitations

- Works only on pages that allow web crawling.
- If a page stops allowing web crawling, The Internet Archive respects this retroactively. In that case the saved page will no longer be accessible. (But the likelihood that it will become inaccessible is significantly smaller than it is for a "live" URL).

Limitations

- Only the page specified is saved, not the rest of the website.
- Embedded content (such as videos) may not be preserved.

It is therefore always a good idea to make your own local copy of the content you need!