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.

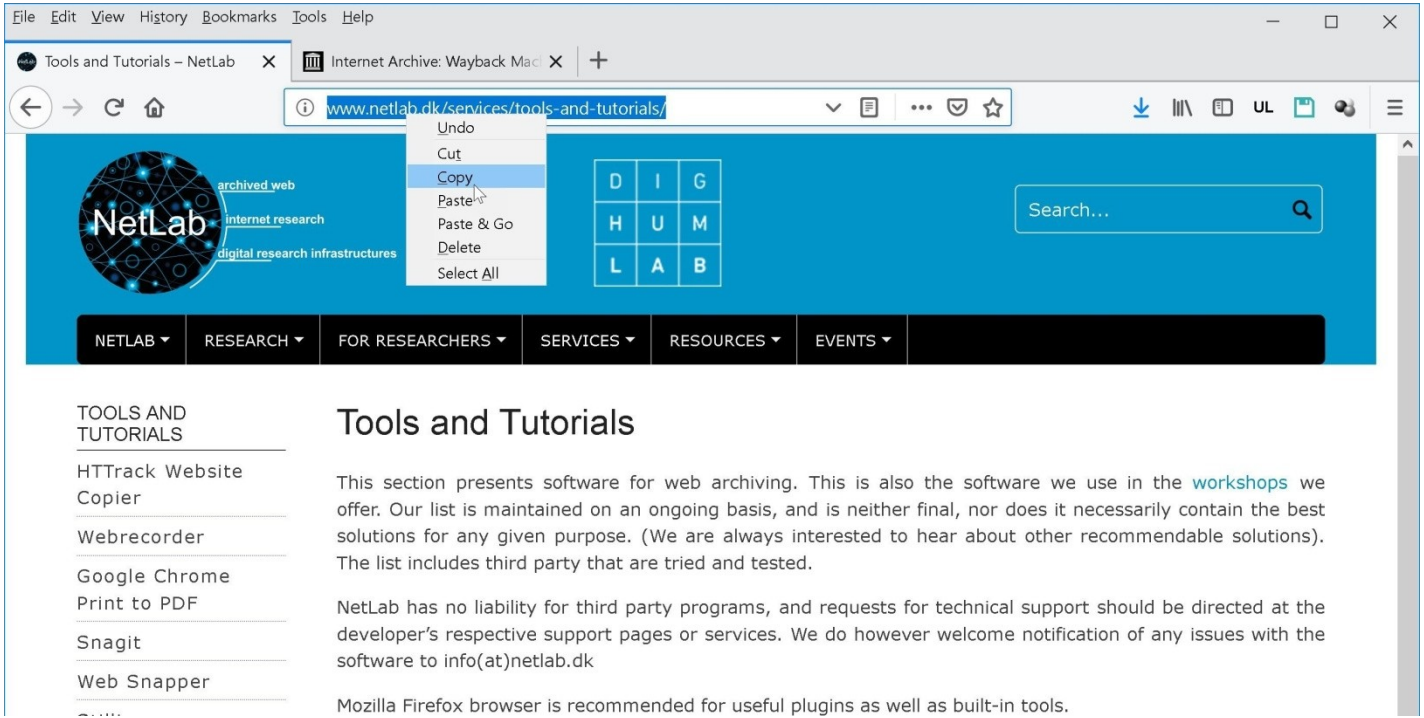
Internet Archive’s ”Save page now” is useful for preservation, but also for references that are more stable than the original URLs of the pages you would like to reference.

Advantages

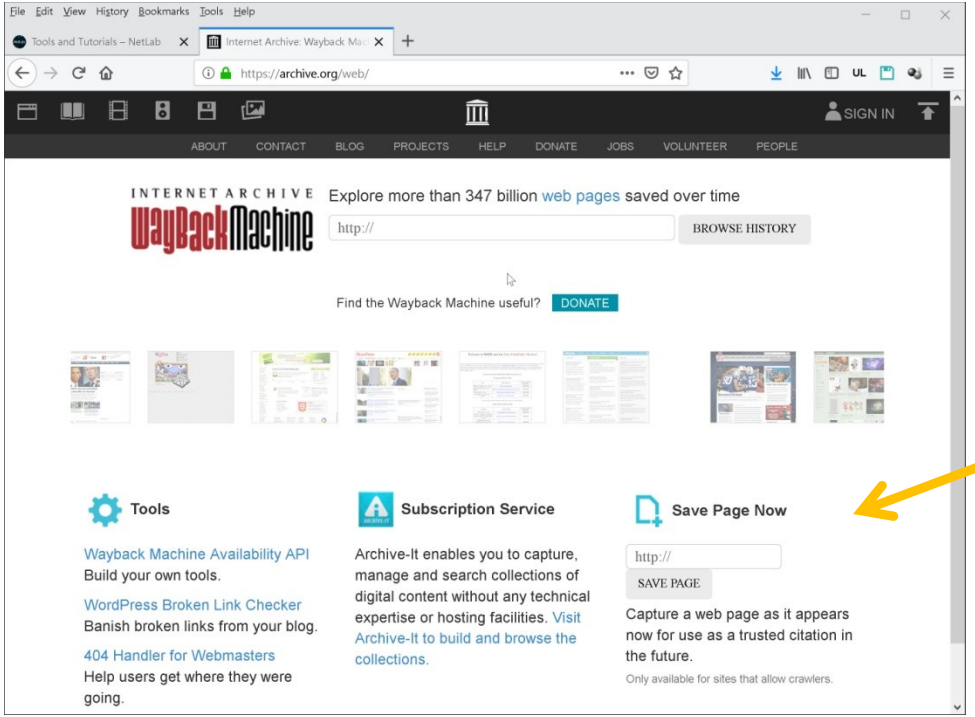
- Easy way to preserve content immediately.
- More stable as a reference than original URL.
- Works with most web pages.

Here is how it works:

1) Copy the URL for the page to preserve



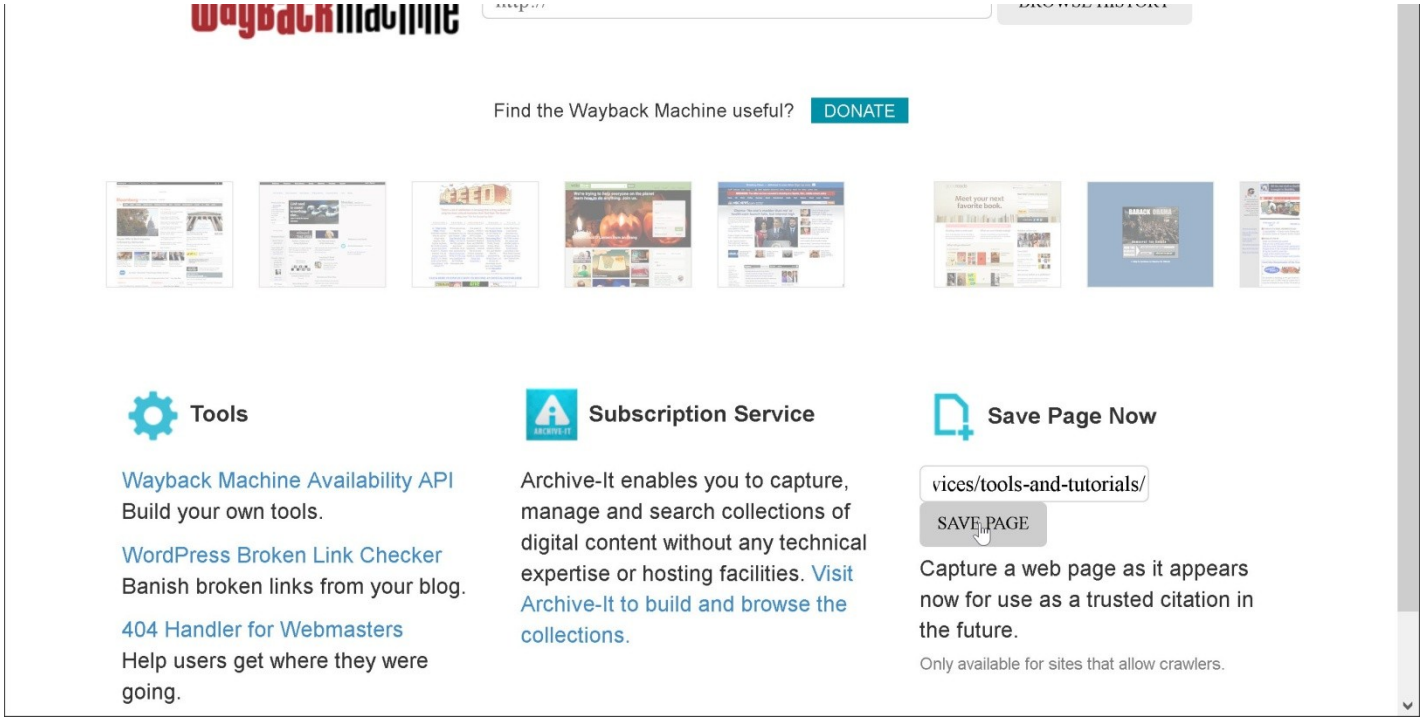
2) Go to archive.org/web



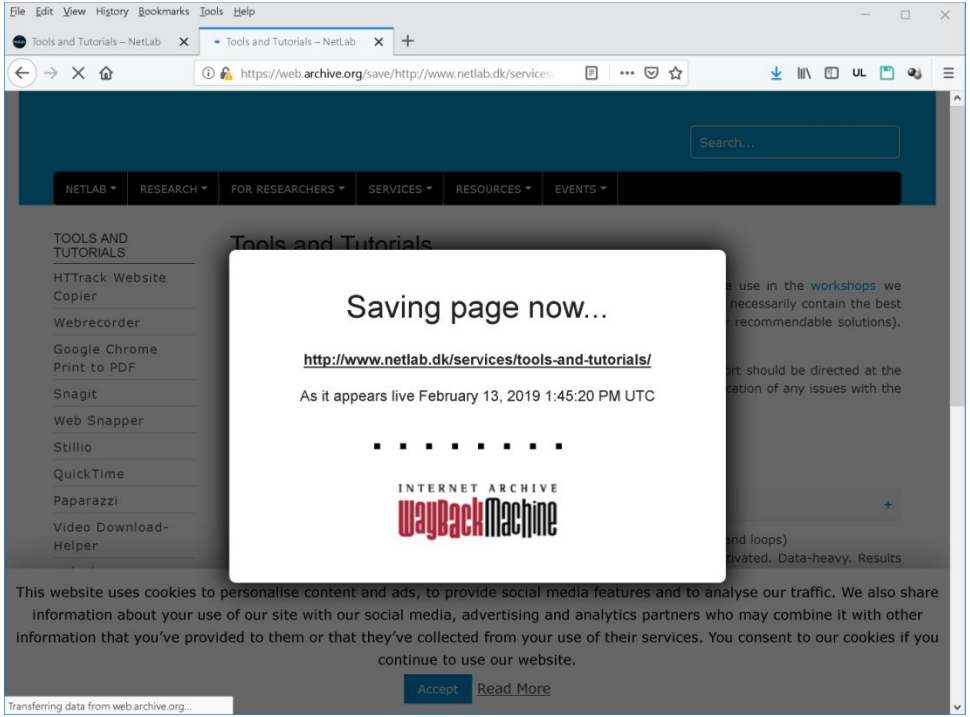
”Save page now” is in the bottom right corner

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

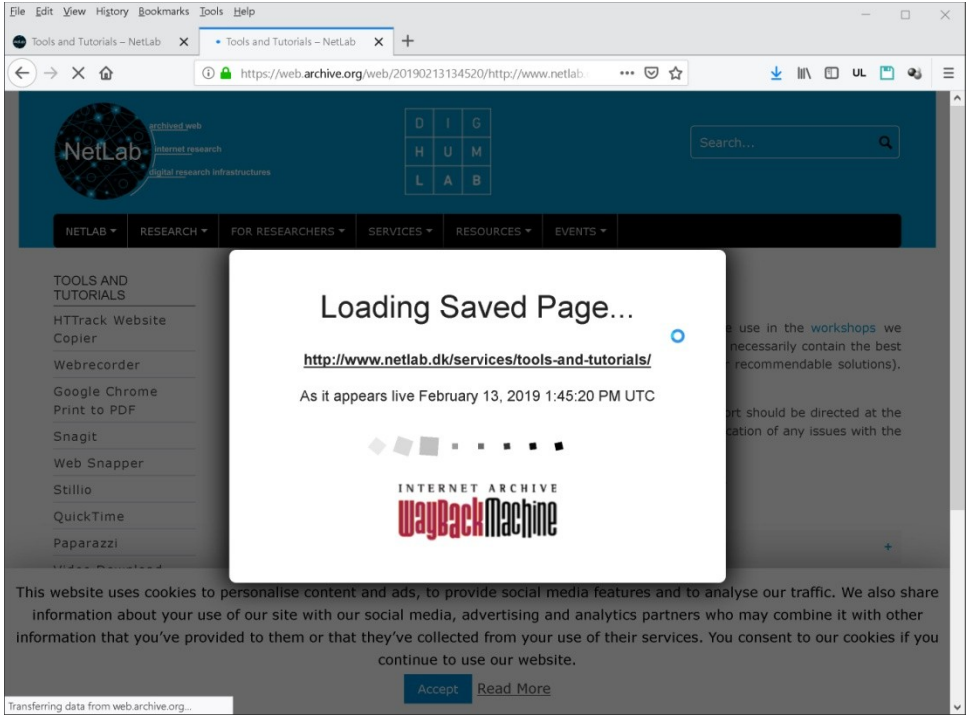
3) Paste URL and click ”Save Page”



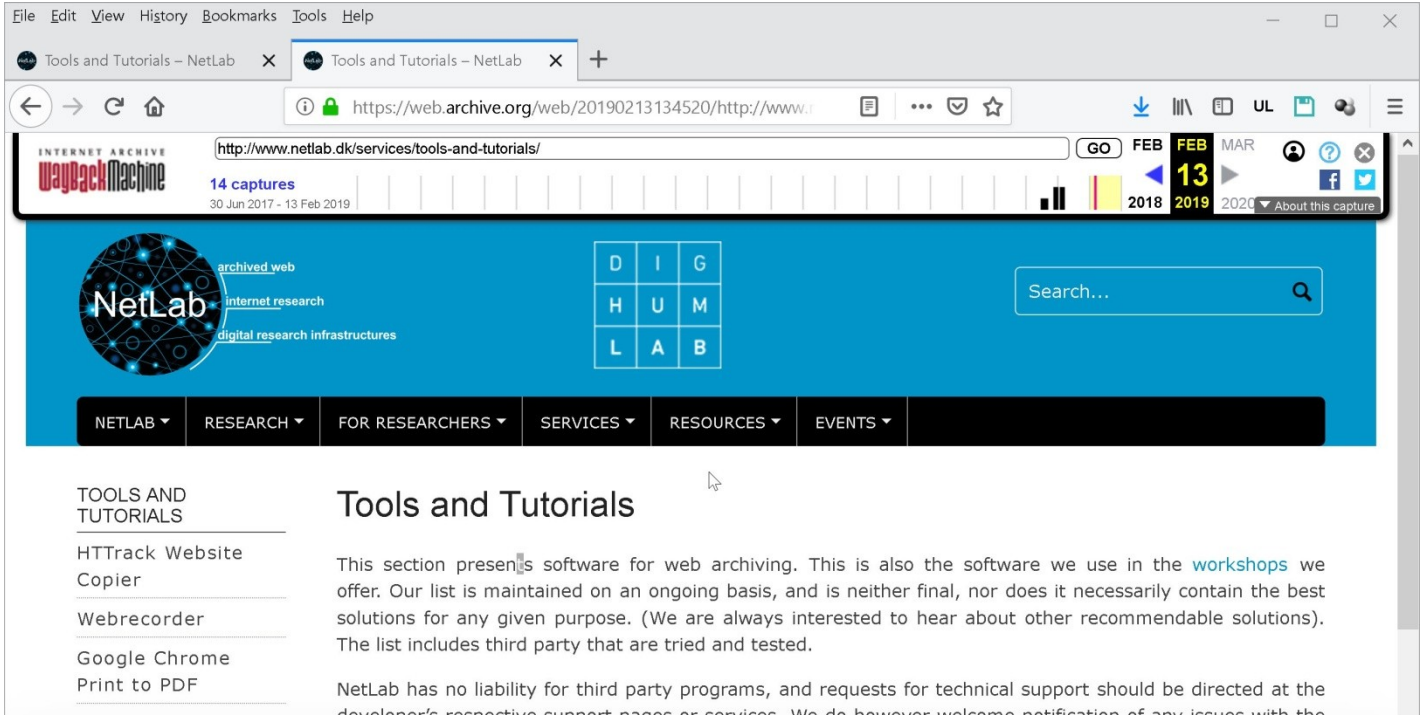
4) The web page is now saving...



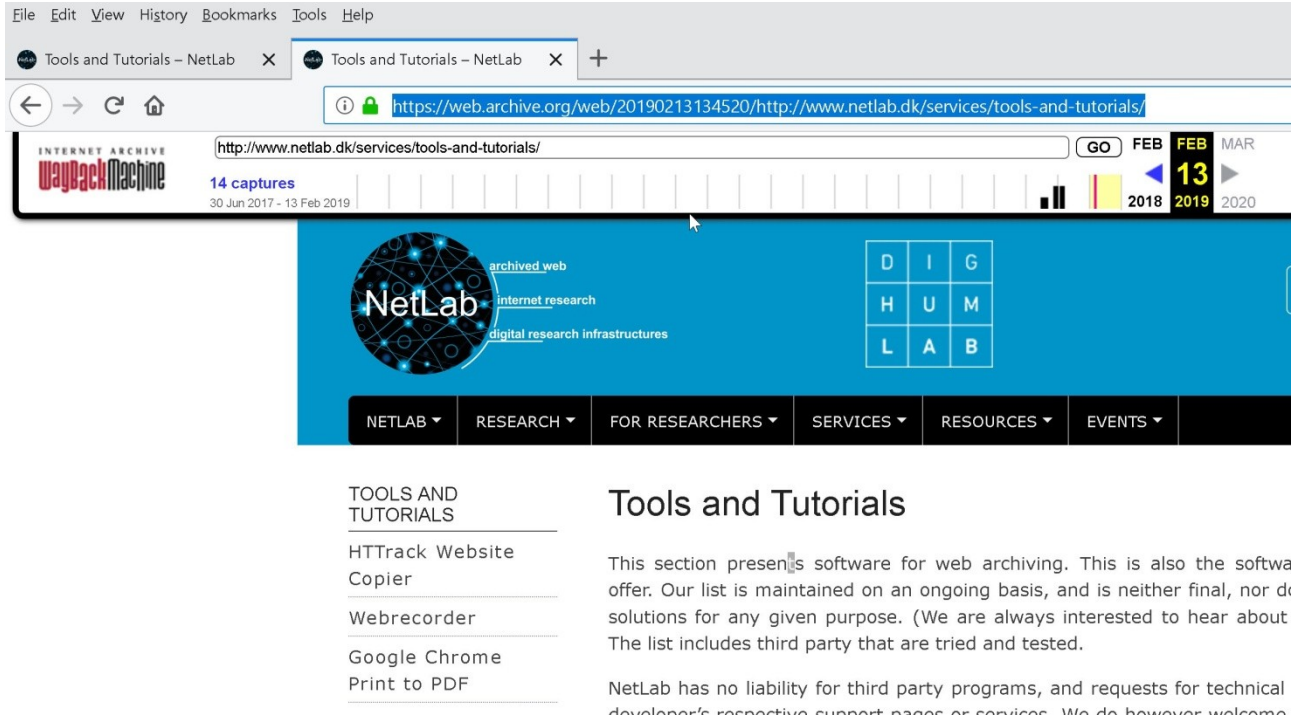
5) After saving the copied page will load:



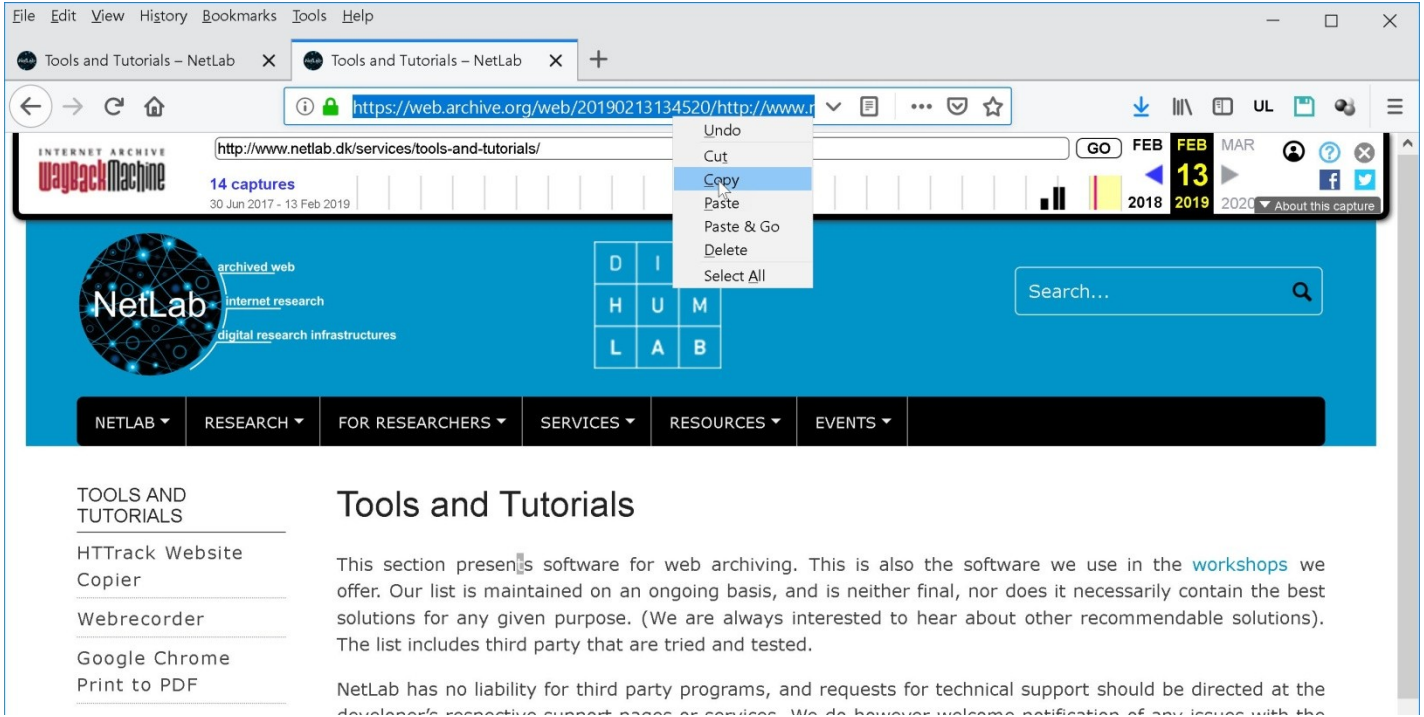
6) Archived copy of the page:



7) Expanded view of full archive URL



8) Copying the archive URL



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.

Archive URL Structure

The second part of the archived URL:

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

20190213134520/ is a time stamp with the formula YYYYMMDDhhmmss. Time of day (last six digits) is GMT.

Archive URL Structure

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 that 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!