



# WEB ARCHIVING IN THE NATIONAL LIBRARY

The Hungarian National Széchényi Library (NSZL) – as national libraries in many countries – has a responsibility for long term preservation of publicly available internet resources, especially websites. By the development framework of the National Library Platform a pilot project has started in spring 2017 in order to establish the professional conditions of a Hungarian web archive.

## WHY?

the Hungarian web should have a past, not just a present

*the history of the virtual world must be researchable, the online imprints of the current events in the real world must be preserved*

the large volume of digital information must be analyzable and visualizable

*there is a need for stable references to online sources in education materials and scientific publications*

## WHAT?

mainly Hungarian websites focusing on culture, science, education or public events

*Hungarian electronic periodicals: journals, magazines, news sites and blogs*

relevant contents related to Hungary on foreign webserver

*publicly available social networking sites of important Hungarian institutions and persons*

## HOW?

thematic collections by topic, genre or institution have crawled periodically

*in some cases written copyright agreements are requested from the content owners to publish the archived copies publicly*

temporal harvests of webpages related important national or world-wide events

*representative crawls of the Hungarian web-sphere once or twice a year*

## About the project: (based on data from May 2023)

IT infrastructure provided  
by KIFÜ and by NSZL

software background: Heritrix,  
Web Curator Tool, HTTrack,  
ArchiveWeb.page, Brozzler, PyWb,  
OpenWayback, SolrWayback...

number of selected websites  
ca. 75,000, total size of the  
archive: 85 TB

the public collection contains  
335 websites, periodicals or  
blogs, and 99 online content  
services of the national library

## Useful links related to web archiving:

NSZL Web Archive's homepage:  
[webarchive.hu](http://webarchive.hu)

Internet Archive: [web.archive.org](http://web.archive.org)  
(the biggest web archive of the  
world: 805 billion archived web  
pages since 1996)

IIPC: [netpreserve.org](http://netpreserve.org) (founded in  
2003, members from around 45  
countries, NSZL is also a member  
since 2018)

[archive.is](http://archive.is) (free on-demand web  
archiving service)



Some sites that have disappeared from the live web but are archived in time