01. Archiving

Purpose

Web archiving preserves web content for future generations and keeps it accessible to the public, even if it is no longer available on the original website.

What is archived and when?

Regular archiving

Websites belonging to the EU institutions, agencies and bodies are archived at least 4 times per year. In general, the scope of regular archiving is limited to websites hosted on the europa.eu domain and subdomains. External websites may be included only if duly justified.

Ad hoc and final archiving

Ad hoc and final archives respond to specific needs at a specific point in time:

- keeping a thematic record or collection of parts of a website (e.g. relating to COVID-19).

- preserving a final record of the content of a website which is to be taken offline or going to change substantially.

Final archives are usually created at the request of the website owner. The archive should be requested at least one month before the date on which the website is to be removed and preferably as soon as the decision to remove it has been made. Requests must be sent to the Web archiving service at the EU Publications Office.

Where to find the archive

The EU web archive is available here: https://archive-it.org/home/euwebarchive

Links to archived content are structured as follows:

https://wayback.archive-it.org/12090/*/URL *of the website you want to consult*

For example:

https://wayback.archive-it.org/12090/*/https://ec.europa.eu/info/index_en takes you to the calendar page, where you can see the different dates on which the English version of the Commission homepage has been captured.

To view the most recent archive, simply look for the last available date.

Web archive content

The Publications Office checks the archives regularly. However, feedback from website owners is extremely useful regarding missing content, whether the archive is displayed in all available languages, or whether sub-sites that have been omitted from the archive.

In archive terminology, archived URLs are called 'seeds'.

All pages with a URL starting with the same root as the seed will be archived.

E.g. for the seed www.webpage.eu/environment/:

- www.webpage.eu/environment/clima will be archived,

- but www.water.webpage.eu or www.webpage.eu/weather are out of scope.

Some types of content are excluded from web archives:

- Databases and some types of dynamic content highly dependent upon human interaction.
  This means that searches will not work, neither will links based on search queries.
- Social media. Some embedded content may appear in the archive. However, do not expect all social media content to be included.
- External links and documents out of scope.
  - The crawl captures all URLs discovered as part of the website, as explained above.

- Any link to a different website (e.g. https://www.un.org/) will not be archived by default.

Preparing sites for archiving

Before revamping or taking all or part of your website offline, you may want to archive it one last time. Prepare your website for archiving by removing all content and files that have no future value (historical, legal, political, research, cultural). Remove any content that is:

- protected by intellectual property rights (e.g. copyright)
- confidential or private
- affected by data protection rules.

Users can navigate archived sites like a live website. However, archiving with a crawler has some technical limitations and as a result certain features may not work, such as:

- the original website's built-in search;
- content that can only be reached after logging in;
- certain navigational elements, e.g. drop-down menus, tick boxes and some maps;
- flash animations and games, streaming media and embedded social media;
- complex JavaScript;
- POST functionality.

How to make a web archiving request

Archiving workflow

1. Regular archiving of living websites

What	How	Who	When
Archiving request	Send an e-mail to OP-WEB-PRESERVATION	Website owner	Upon establishment of a new EC and/or DG website
Analysis of request		OP WP team	
Approval/rejection of request	Email with justification of conclusions to website owner and Comm Europa Management	OP WP team	
For accepted requests			
Regular crawling	Remote crawling	OP WP team	At least four times per year
Quality control	Visual/manual check of quality of the crawl, and feedback to OP WP team	Website owner	Upon invitation, sent by OP WP team, or any time.
Patching	If needed and if possible: improving the quality of the archived version	OP WP team	Upon reception of website owner's feedback on quality or as a part of the regular quality control
Acceptance /rejection of crawl	Email to OP-WEB-PRESERVATION	Website owner	
Publication /takedown of crawl		OP WP team	

2. Ad hoc archiving of websites that are to be taken offline or changed substantially

What	How	Who	When
Archiving request	Send an e-mail to OP-WEB-PRESERVATION	Website owner	At least 1 month before the site will be taken offline /changed
Analysis of request		OP WP team	
Approval/rejection of request	Email with justification of conclusions to website owner and Comm Europa Management	OP WP team	Maximum 1 week after reception of CEM approval
For accepted requests			
Planning	Discussion of deadlines and crawl specifications	OP WP team and website owner	Upon approval of the request
Crawling	Archiving following crawl specifications	OP WP team	According to planning agreed with website owner
Quality control	Regular quality control by OP WP team	OP WP team	Upon invitation, sent by OP WP team
	Visual check of quality of the crawl, and feedback to OP WP team	Website owner	

Patching	If needed and if possible: improving the quality of the archived version	OP WP team	Upon reception of WO feedback on quality or as a part of the regular quality control
Acceptance	Email to OP-WEB-PRESERVATION	Website owner	
Publication		OP WP team	
Redirections (if desired)	See our section on redirections	Website owner	

Takedown policy

Under certain circumstances, it may be necessary to hide pages in the web archive from public view.

Anyone can submit a motivated takedown request via email to OP-WEB-PRESERVATION.

Takedown will be considered only if the page:

- includes one of the following types of content:
  - personal or sensitive personal information, as defined by Regulation (EU) 2018/1725 on the protection of personal data as processed by EU institutions, bodies, offices and agencies
  - copyright protected material for which the necessary rights are not held
  - defamatory or obscene material or messages
  - content which may cause serious and real administrative difficulties to the website owner
- was published in good faith, but due to a change in circumstances its takedown is now considered appropriate
- was published in error and takedown is necessary to correct the mistake.

Legal information

Copyright

© European Union, 2019

The Publications Office carries out web archiving to preserve the EU websites. Most of the archived content of websites in the EU web archive (EUWA), is under EU (or EU institutions, agencies or bodies) copyright. Ownership and copyright of websites in the EUWA remain the responsibility of the website owners.

Unless otherwise stated, material obtained from the EUWA may be freely reproduced. This general principle can be subject to conditions, which may be specified in individual copyright notices. It does not apply to photographs, videos, pieces of music or other material subject to intellectual property rights of third parties (non-EU). In such cases, permission to use the material must be sought directly from the copyright holders. The Publications Office does not guarantee that all third-party content is appropriately marked.

All logos and trademarks are excluded from the abovementioned permission.

Any queries regarding the above should be addressed by email to OP-COPYRIGHT@publications.europa.eu

 See also the Privacy statement.

Contact and support

Need further assistance on this topic? Please contact the team in charge of the Europa Domain Management (EU Login required).