

WEB ARCHIVING

Use this quick guide to understand how to capture and save copies of web content, so it can still be accessed even if the original changes or disappears.

IN A NUTSHELL: You may want to include web-based content in your thesis, either created by you or drawn from external sources. However, web content does not last indefinitely and can become inaccessible very quickly. Even if it's your content hosted on your website, it may not be realistic to maintain or sustain access to it over the long term, making it difficult to rely on in the future. Web archiving tools offer a practical way to capture and preserve web content, helping ensure it remains accessible and understandable over time.



THINGS TO KNOW

- **Websites are not stable by default.**

Websites depend on servers, software, and ongoing maintenance to remain available. Content can change frequently as part of normal updates or be removed altogether if a site is no longer maintained or supported. As websites are directly connected to modern life and contemporary organizations, they tend to change in line with ongoing shifts in business, government, society, and so on. This means web content may not remain the same—or remain available—over time, making it difficult to depend on as a long-term record.

- **Web archiving creates a snapshot in time.**

Web archiving tools capture a version of a webpage as it appears at a specific moment. This creates a “snapshot” that can be accessed later, even if the live site changes or disappears. The archived copy is fixed in time and does not update, providing a stable record of the content as it was when captured.

- **Not all content can be captured fully.**

Some content—such as interactive features, databases, or material behind logins—may not be captured completely by web archiving tools. Archived versions may not behave, be ‘replayed,’ or made available exactly like the live site.

- **Design choices affect how well content can be archived.**

If you are creating web content, be aware that content which is contained in a standard web page (such as text and images with a clear structure and a stable link) is more likely to be captured successfully by web archiving tools. Embedded content and content that relies heavily on scripts, dynamic features, search, or external services is harder to archive reliably.

- **You can archive both your own and others’ content.**

Web archiving tools can be used to capture your own web-based research outputs as well as important third-party content that you cite. In most cases, publicly accessible web content can be archived, but some material may be restricted (for example, content behind logins, paywalls, or subject to legal or technical limitations). Note that audiovisual content, such as YouTube, is challenging to capture and replay.



THINGS TO DO

Design web content with archiving in mind.

If you are creating web-based content, use simple, well-structured pages where possible. Avoid relying heavily on dynamic features (e.g. content that only loads through interaction), or platform-dependent features. Where more complex features are essential to your work (such as interactive elements or custom-built applications), be aware that these may not be fully captured by web archiving tools.

Create an archived version of important web content.

For web content (your own or third-party) that is central to your research, create an archived version so that it can still be accessed if the live site changes or disappears. See below for tools to help you do this.

Check what has been captured.

After capturing a web archive version of a page, review it to make sure key content is present and works as expected. If anything important is missing, document this and consider how best to represent it—for example, by briefly describing what is not available in the archived version or providing additional context where needed. The level of detail should reflect how important the missing content is to your work.

Include links to archived content alongside original links.

When citing web content from your thesis, include both the original URL and the archived version where possible. This preserves the original context of the source while also providing a stable, time-stamped copy that can still be accessed if the live page changes or disappears.

Keep copies of critical web content where appropriate.

For web content that is central to your research, consider keeping additional copies alongside your thesis (where copyright allows), rather than relying on a link or archived version alone. This might include saving screenshots of especially pertinent pages, downloading files, creating transcripts of audio or video, or exporting key content. Using more than one method helps ensure the material remains accessible and understandable, even if links break or archived versions are incomplete.



WEB ARCHIVING TOOLS AND RESOURCES

- **Internet Archive “Save Page Now”**

The Internet Archive provides a [Save Page Now](#) feature via the Wayback Machine. This allows you to capture a snapshot of a single webpage and generate a stable archived link. This approach is best suited to individual pages rather than entire websites or complex web resources. You can use it when:

- You want to preserve a specific page at a specific moment
- You are citing web content that may change or disappear
- You are archiving your own web-based work at the page-level

- **ArchiveWeb.Page**

For more complex web content, consider using [ArchiveWeb.Page](#). Unlike the Internet Archive’s Save Page Now feature, this tool allows you to capture multiple related pages as part of a single web resource. It also gives you more control over the archived content—you keep the copies and can decide how and whether to share them. In some cases, it may work better for capturing dynamic content (such as interactive features or social media) too, but may take a little more time to get familiar with.

- **Ask about archiving larger or more complex websites.**

If your thesis includes a substantial website (for example, a standalone project, digital exhibition, or online research output), speak to your university library or archives team. They may be able to capture your site as part of a broader web archiving programme (if this service is available at your institution). This is particularly useful for sites that are too large or complex to archive effectively using freely available tools.

