MANUAL ON PRACTICES OF ONLINE ARCHIVING AND CONSERVATION EVIDENCE.

The Internet is not an archive! The average lifespan of a web page is 92 days.

Some web archiving practices exist which allow you to preserve the history of a page and its metadata. This process is not automated and **you must manually archive the pages you whish to save.**

Through the Wayback Machine tool, a user can :

- Archive a page in one click
- Go back through the history of archived pages and access previously archived versions of the page.

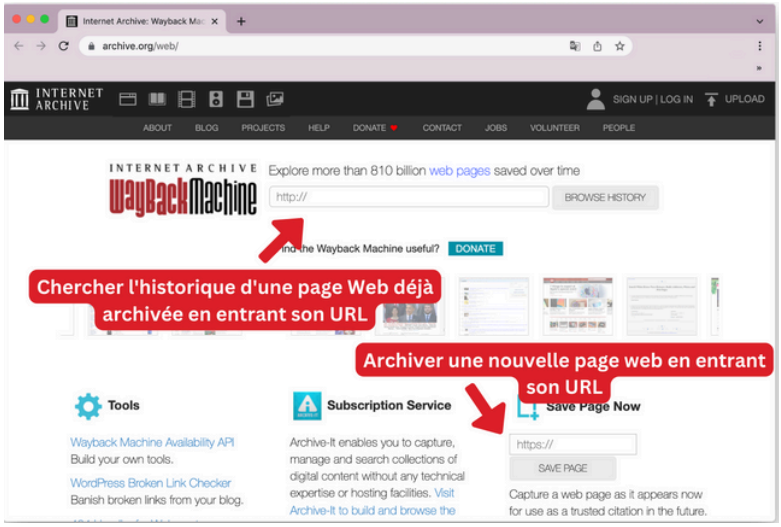
Even if the URL directs the user to a page that no longer exists (see appendix 1), Wayback is able to recover the information from this page if it has been archived. However archiving cannot retrospectively give you acess to the information attached to a dead link.

Archiving : How to use Wayback Machine

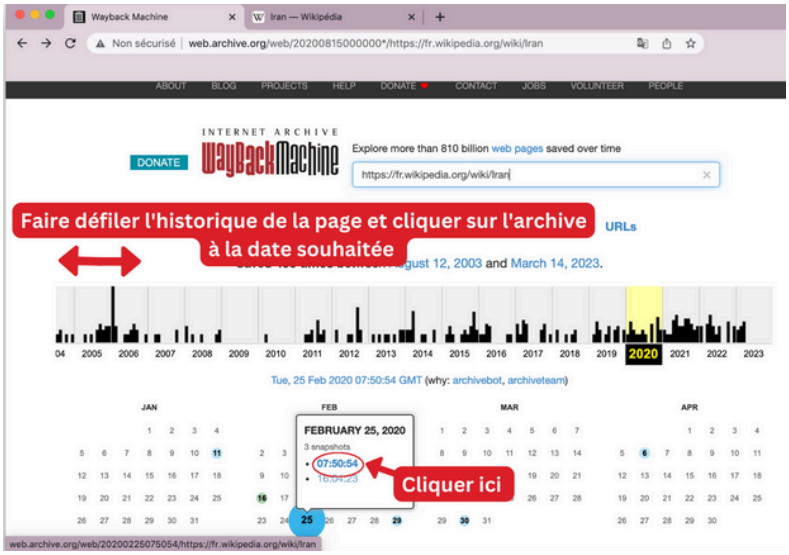
http://web.archive.org/

1) Through the search bar, check if the URL has already been archived. If the URL does not exist enter the new URL under the "Save Page Now" section.

Warning: To archive a video or photo on Twitter, ensure that the account settings allow the viewing of sensitive images.



There are other very efficient archiving websites such as : https://archive.ph/



2) Once the URL is saved, the page history will be displayed and will allow you to view the modifications made to the page with precise dates and times of modifications.

3) By clicking on a specific date, you can view the page as it was captured. You can also use the timeline bar to scroll through a web page and observe its evolution.

Storing on devices

Safe storing on devices

It is best not to store important evidence on devices connected to the cloud. If an information needs to be shared across the internet, **it is best to do it through encrypted clouds and email addresses such as proton** [see: https://proton.me/]. Proton unlike Google and Outlook encrypts information even before saving it to the cloud, which means that the server does not have access to the stored information.

Warning : If you use a cloud, you must always try to save the information on a computer and **make regular backups on a hard drive** so as not to lose the information. It is also preferable to **have a dedicated email address, password and account for information storage websites online** (try not to use personal adresses !).

It is sometimes impossible to download the videos, in this type of situation you can use Excel to record the URLs of the evidence and name them (don't forget to archive these pages!). As a reminder, Microsoft Excel and Google Sheet, which are the two most used online platforms for creating Excel files, are not encrypted platforms. You must therefore be careful to **limit the sharing options**, the most preferable option remains to **use the local excel app on your device** and if needed save a copy in an encrypted cloud.

> **Warning :** When working on **public WiFi networks**, activity can easily be spied on. In particular public wifi at train stations and airports. **Get a VPN** to protect your online activity and passwords!

On a phone

It is possible to download videos and information from a phone. Not all software under Apple iOS and Google Android offers the ability to name a video in the phone gallery. It is therefore better to store the evidence and videos on a drive application [proton is accessible on Apple and Android] in order to keep the information there and name it correctly.

Preserving the evidence

The more a video is manipulated, the more complex its metadata becomes. If it is necessary to transfer the video and analyze it (pass it from device to device) or change its format (switch from mp4 to MOV), all these actions can be detrimental to the admissibility of the evidence within the legal framework. **The Berkeley protocol therefore recommends copying the video in order to analyze it and keeping the original version intact.** This is true of all files (vidéo, image, text file, etc.)

Naming

Archiving evidence is not only about preserving it, it is also about ensuring that it can be easily cited, so it is vital to label the evidence:

yyyy-mm-dd_CreatorName_IncidentDescription.

During a collection, you must systematically :

- Do a screenshot of the proof with the date and time visible, as well as the information of the publisher (ex: user name)
- Keep the URL of the page where the information is located and archive it.
- If multimedia content is included in the evidence download this content separately by naming and storing it correctly and securely.

Citing websites and online files

In research

The Internet is not an archive! To rely on a web page and the information it contains for research, the information and the web page must remain identical so as not to change the research material.

In law

According to the International Criminal Court, the admissibility of audiovisual evidence taken from the internet (open source evidence) is based on three factors. The relevance of the evidence, the probative value of the evidence and the absence of prejudicial effect. As explained in the Rome Statutes on the Handling of Evidence, **open source evidence must be subject to metadata evaluation for veracity to be established**.

Any case based on open source evidence such as audiovisual evidence is at risk if the online information is modified, damaged or corrupted.

> **Warning:** Videos that are blurred or have subtitles are considered corrupt. Effort should be made to find the source of the original video.

Open Source information collection form

In order for the metadata of an Open Source proof to be collected, the conditions in which the proof was collected are important. The form proposed by the **Berkeley Protocol** (see annex 2) stipulates that the date and time of collection of the information, as well as the identity of the individual collecting the evidence and the IP address of the device on which the evidence has been collected must be systematically stipulated.

See “Online data collection form” sample in the Berkeley Protocol

Annex p.86

ANNEX IV

Annex IV

Online data collection form	
1. Collector information	
Investigation:	
Collector:	
Collector IP address:	
Start of collection (date/time stamp):	
End of collection (date/time stamp):	
2. Target information	
Web address (URL):	
HTML source code:	
Screenshot:	
Captured data:	
IP address(es):	
3. Collection package information	
Collection package file name:	
Collection package hash list:	
Hash of collection package hash list file:	
4. Services used	
Software product(s):	
Time service:	
IP service:	
WHOIS service:	

Annexe 1 :

Web Pages

Construction of a web page

A web page is a medium through which files or information are made accessible. A web page is accessible through a URL and includes **metadata**.

URL or “Uniform Resource Locator” is a web address that allows you to navigate in order to find and access information on the internet. A URL does not guarantee the quality of the information and is only a citing system which enables you access.

Metadata: Metadata is the information attached to online information/file. If a letter is data, the letter's metadata is the date sent, the stamp, the weight of the letter, and the sender's address. On the internet all this information can also be traced.

A web page in time

A web page can deteriorate over time to the point of no longer containing the same information. The average lifespan of a web page is 92 days. So a URL can lead users to a page that no longer contains the same information.

The information may have been moved, the server architecture may have been modified, or the file and online information may have been deliberately targeted and damaged. In certain cases viruses can attack information in a targeted manner.

Annexe 2 :

Berkeley protocole

With the growing use of social networks as a tool for information and activism , the internet has become an eminently political place that allows the propagation of information, photos and videos. This information can be useful in a legal framework and used as evidence, it is in this regard that the United Nations published in 2022 the Berkeley Protocol on Open Source Digital Research.

This protocol seeks to standardize online information collection practices so that they can be used by the IPC as well as other international bodies. This manual complies with the protocol's recommendations.

Bibliography :

Alexa Koenig et Mehandru Nikita. « Open Source Evidence and the International Criminal Court », 2019. https://harvardhrj.com/2019/04/open-source-evidence-and-the-international-criminal-court/.

Neal, James G. « The Integrity of Research Is at Risk: Capturing and Preserving Web Sites and Web Documents and the Implications for Resource Sharing ». IFLA, 3 juillet 2014, 24.

Nielsen, Janne. « Using Web Archives in Research – an Introduction ». NetLab, 2015.

Peyssard, Jean-Christophe. « Archivage des données du web : sauvegarde et citabilité pour la recherche ». CNRS, juillet 2022. https://distam.hypotheses.org/3931

Teszelszky, Kees. « Introduction: Digital Humanities and the Use of Web Archives ». International Journal of Digital Humanities 2, no 1-3 (novembre 2021): 1-4. https://doi.org/10.1007/s42803-021-00040-5.