CONSERVATION CENTER

for Art & Historic Artifacts

Web Archiving
with Kaitlyn Pettengill

Conservation Center for Art & Historic Artifacts

In a typical year:

- Preservation Services specialists complete over 50 survey projects.

- Approximately 60 CCAHA-sponsored workshops, conferences, webinars, and training sessions are presented.

- The Digital Imaging Services staff digitizes thousands of pages of fragile archival documents, books, and photographs.

- Conservators assess and treat more than 6,000 individual artifacts, from over 400 clients.

- Housing & Framing Services house approx. 75% of the artifacts treated (folder, sleeve, box, mat and frame, or sealed package).

YOUR WEB CONTENT IS WORTH PRESERVING

AGENDA

- History of web archiving

- How web archiving works

- Determining a project scope

- Creating an archivable site

- Selecting tools

WHAT IS A WEB ARCHIVE?

DEFINITIONS OF WEB ARCHIVING

- Not an archive ON the web, but an archive OF the web.

- "The process of collecting, preserving, and providing enduring access to web content" (Society of American Archivists)

- "A series of steps that work together for an end goal: to interact with a website as it looked on the day that it was archived" (Internet Archive)

Illustration courtesy Chester County Historical Society

The Conservation Center for Art and Historic Artifacts (CCAHA), established in 1977, is one of the largest regional conservation laboratories in the country. A nonprofit regional conservation laboratory, CCAHA serves cultural, educational, and research institutions; individuals; corporations and private organizations throughout the United States.

CCAHA specializes in the treatment of art and historic artifacts on paper. Treatment services are available for any work of art or historic artifact on paper, such as drawings, prints, maps, posters, historic wallpaper, architectural drawings, musical scores, photographs, rare books, scrapbooks, manuscripts, and related materials such as parchment and papyrus.

CCAHA also provides surveys to assist organizations with preservation planning; workshops and seminars; internships; and emergency assistance.
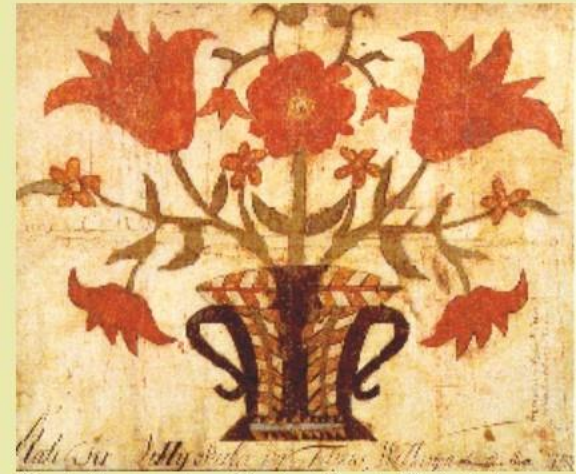
C

- Quality work done by experienced professionals
- Specialization and expertise in conservation of paper-based works of art and artifacts
- Ability to treat oversize materials, complete large volume of work, and meet client deadlines
- Full range of preservation services including treatment, on-site consultation, and emergency assistance
- Newly expanded state-of-the-art laboratory with 24 hour monitored security systems
- Twenty years of successul service and experience

Membership Information	Operational Procedures	Conservation Treatment	Internet Resources
	News & Events	Consultations and Educational Programs	

The
Conservation Center
for Art and Historic Artifacts

• About Our Organization
• Services
• Membership Information
• Publications
• Workshops
• CCAHA News
• Disaster Services
• Meet Our Staff
• Internet Links

The Conservation Center for Art and Historic Artifacts (CCAHA), established in 1977, is one of the largest non-profit regional conservation laboratories in the country. CCAHA serves cultural, educational, and research institutions; individuals; corporations; and private organizations throughout the United States.

CCAHA specializes in the treatment of art and historic artifacts on paper. Treatment services are available for drawings, prints, maps, posters, historic wallpaper, architectural drawings, musical scores, photographs, rare books, scrapbooks, manuscripts, and related materials such as parchment and papyrus.

CCAHA's organizational structure includes a board of directors comprised of representatives of cultural institutions and leaders in business and in the community; an executive director; and an expertly trained staff of professionals. CCAHA provides:

▸ Quality work done by experienced professionals who specialize in the conservation of paper-based art and artifacts
▸ Full range of services including treatment, on-site consultation, surveys, preservation planning, workshops, internships, and emergency assistance
▸ State-of-the-art laboratory with 24-hour monitored security systems
▸ Ability to treat oversize materials, to handle a large volume of work, and meet deadlines
▸ Over twenty years of successful service and experience

Audubon print courtesy of Lehigh University.

Upcoming Workshops with CCAHA
click on the icons for details

Water, Water Everywhere: Recovering Water-Damaged Paper Collections
October 12, 2000

To Outwit Time: Preserving Brittle Paper
October 27, 2000

To Outwit Time: Preserving Brittle Paper
November 8, 2000

New! On-line Technical Bulletins
click on the icons for details

Managing a Mold Invasion: Guidelines for Disaster Response
Technical Series No. 1

CONSERVATION CENTER
for Art and Historic Artifacts

The Conservation Center specializes in the treatment of art and historic artifacts on paper. Established in 1977, CCAHA is one of the largest non-profit regional conservation labs in the country.

Home | About the Center | Services | Education | Funding Opportunities | Publications | Join & Support the Center | Career Opportunities | Contact

Photograph by Jessica Hoffman

Crowds Flock to CCAHA

CCAHA opened its doors on January 19 for an open house and book signing with Audubon biographer, Richard Rhodes.

Guests were treated to individual discussions about a range of artifacts being treated at CCAHA—from the first printed map of Philadelphia to the secret journals of the 1774 Congress to Wright Brothers airplane photographs.

Photo by Jessica Hoffman

Lafayette College Papyrus Gets New Life

Conservator Soyeon Choi examines part of a Egyptian papyrus believed to date from the first century C.E.

>> See treatment

Read the newest *Art-i-facts*

The new issue of *Art-i-facts* is packed with interviews, articles, and tips you won't want to miss. Click here for more infomation

Check out CCAHA's new workshops!

Our 2005 Collections Care Workshops have just been posted. Register now by clicking here.

Subsidized surveys are available!

Download the details by visiting our Surveys page by clicking here.

search our site

WHY IS WEB ARCHIVING IMPORTANT?

le.com

LOSS HAPPENS QUICKLY

- 25% of web pages posted between 2013 and 2023 (Pew)

- 8% of sites published in 2023 were gone in less than a year (Pew)

- Typical web page lasts 90 to 100 days before being changed, moving, or disappearing (Old Dominion University)

- 10% of event related content posted to social media platforms is lost after one year (Old Dominion University)

We're losing our digital history. Can the Internet Archive save it?

16 September 2024

Share   Save +

Chris Stokel-Walker

(Credit: Serenity Strull/ Getty Images)

HOW DOES WEB ARCHIVING WORK?

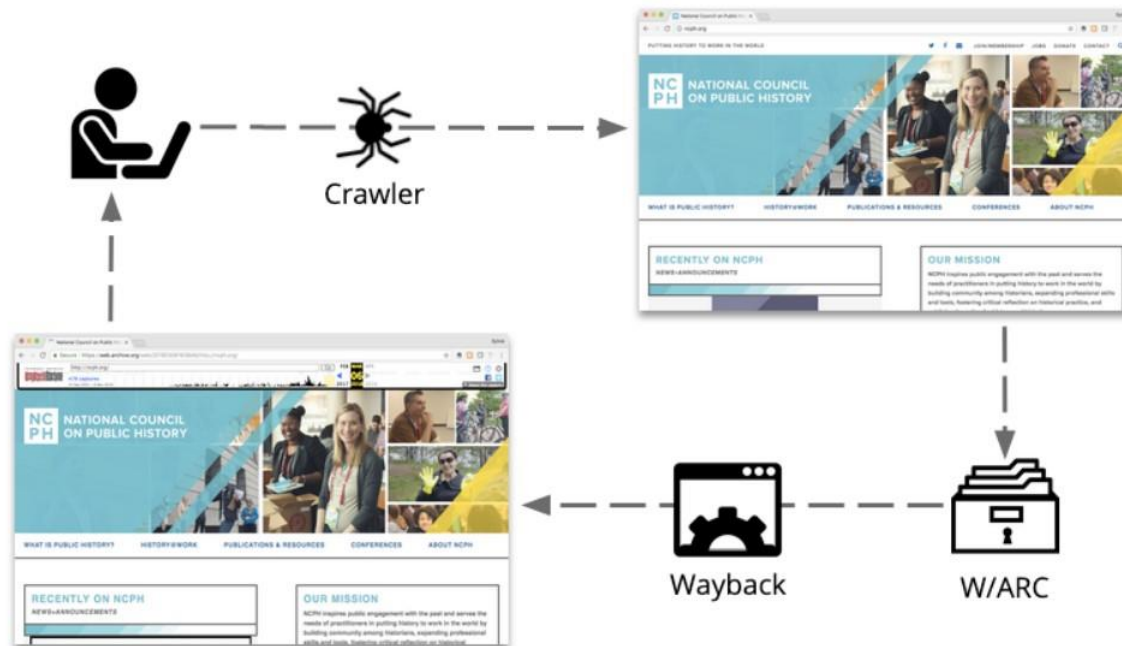# WEB CRAWLERS

- Also be called a harvester, robot, or spider

- A crawler is software that automatically traverses the web

- Can capture HTML code, JavaScript, images, video, metadata

WEB CRAWLING PROCESS / WORKFLOW

- Identify a site

- Set up a bot crawl

- Save as a **W**eb **ARC**hive (or WARC) file

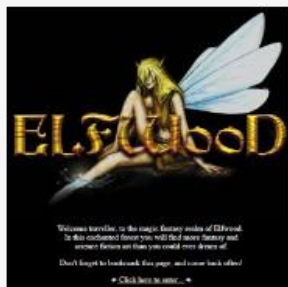- Ingest into the archive

- Go live via the playback tool

COLLECTION
Web Cultures Web Archive

About this Collection   Collection Items

➖ **Featured Content**

Cute Overload! ;)

Elfwood - Science Fiction and Fantasy Art, FanArt & Stories

Urban Dictionary: Define Your World

CREEPYPASTA.COM - Scary Paranormal Stories & Short Horror ...

Internet Meme Database | Know Your Meme

● ○ ○

About this Collection

Rights & Access

🔊 Listen to this page ▶

Expert Resources

Data Mining Memes in the Digital Culture Web Archive

Web Archive Datasets

About this Collection

The Web Cultures Web Archive includes sites documenting the creation and sharing of emergent cultural traditions on the web. The mission of the American Folklife Center is to document traditional cultural forms and practices, and the proliferation of smart phones, tablets, and wireless Internet connections has positioned networked communication as a space where people increasingly develop and share folklore. This collection, co-curated with scholars who study digital culture, captures a set of websites that

WHERE DO I START?

DETERMINING SCOPE

- What is the overall project?

- Why do you want to archive a particular site or page?

- What elements of the site are the most important?

ARTIFACTUAL VERSUS INFORMATIONAL VALUE

Artifactual or intrinsic value:

- The aesthetic experience produced by the original work
- The construction methods, such as the binding

Informational value:

- The "words in the text" rather than the text itself
- Data that can be extracted

Pierpont Morgan Library. Acts of the Apostles [MS G.67]. 5th Century.

NORTHANGER ABBEY

A NOVEL

BY JANE AUSTEN

NEW EDITION

NORTHANGER ABBEY.

CHAPTER I.

NO one who had ever seen Catherine Morland in her infancy, would have supposed her born to be a heroine. Her situation in life, the character of her father and mother, her own person and disposition, were all equally against her. Her father was a clergyman, without being neglected, or poor, and a very respectable man, though his name was Richard—and he had never been handsome. He had a considerable independence, besides two good livings—and he was not in the least addicted to locking up his daughters. Her mother was a woman of useful, plain sense, with a good temper, and, what is more remarkable, with a good constitution. She had three sons before Catherine was born; and, instead of dying in bringing the latter into the world, as any body might expect, she still lived on—lived to have six children more—to see them growing up around her, and to enjoy excellent health herself. A family of ten children will be always called a fine family, where there are heads and arms and legs enough for the number; but the Morlands had little other right to the word, for they were in general very plain, and Catherine, for many years of her life, as plain as any. She had a thin awkward figure, a sallow skin without colour, dark, lank hair, and strong features;—so much for her person;—and not less unpropitious

B
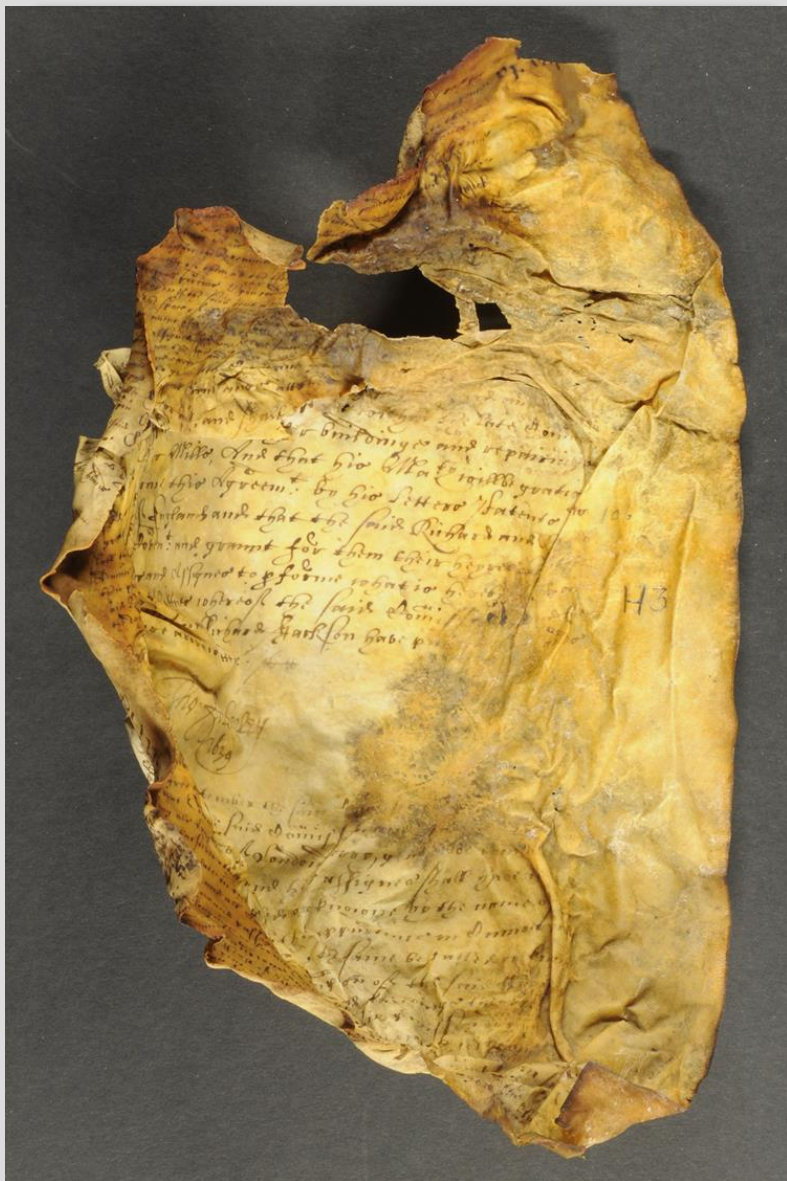
one and twenty years from the feast of Phillip and Jacob *now last past*, for *and under the yearly rent* of six pounds thirteen shillings four pence *sterling* payable to his Majesty, his heirs and successors, into the *rece*ipt of his Majesty's Exchequer in England, or unto the hands of *th*e Receiver of his Majesty, his heirs and successors for the time being for the revenue of Londonderry, *also* Derry, and Coleraine or either of them at the feasts of All Saints and Phillip and Jacob by equal portions. Upon condition that the said term shall cease *and* be void if the said rent or any part thereof be behind and un*paid* by the space of fifty

London Archives. Great Parchment Book of the Honourable The Irish Society. Late 1630s.

ARTIFACTUAL VERSUS INFORMATIONAL VALUE

Artifactual or intrinsic value:

- The aesthetic experience produced by the original work
- The construction methods, such as the binding

Informational value:

- The "words in the text" rather than the text itself
- Data that can be extracted

ARCHIVING TIPS FROM LIBRARY OF CONGRESS

- ▪ *Identify* where you have websites

- ▪ *Decide* which information has long-term value

- ▪ *Export* the selected information

- ▪ *Organize* the information

- ▪ *Make* copies and *manage* them in different places

IDENTIFY WHERE YOU HAVE WEBSITES

- Locate all your content on the Web, including personal websites and social media sites and services

- Be sure to include current information as well as any older (archived) content

DECIDE WHICH INFORMATION HAS LONG-TERM VALUE

- Pick the information that you feel is especially important

- You can pick individual pieces of information or you can select whole pages or websites

EXPORT THE SELECTED INFORMATION

- If saving a limited amount of information, you can use the "save as" command to export as individual files

- If saving entire pages or sites, check into automatically exporting them as a series of linked files

- Save metadata for your Web content information, such as site name or date created

ORGANIZE THE INFORMATION

- Give individual files descriptive file names

- Create a directory/folder structure on your computer to put the saved information

- Write a brief summary of the directory structure and its files

MAKE COPIES AND *MANAGE* THEM IN DIFFERENT PLACES

- Make at least two copies of your selected information—more copies are better!

- One copy can stay on your computer or laptop; put other copies on separate media

- Store copies in different locations that are as physically far apart as practical

HOW DO I MAKE SURE THAT MY WEBSITE IS ARCHIVABLE?

CREATING PRESERVABLE WEBSITES

- Follow web standards and accessibility guidelines

- Be careful with robots.txt exclusions

- Use a site map, transparent links, and contiguous navigation

- Maintain stable URIs and redirect when necessary

- Consider using a Creative Commons license

- Use sustainable data formats

- Embed metadata

Is your website Archive Ready?

http://www.yoursite.com/   Check now »

What is ArchiveReady?

An online tool which evaluates **if a website will be archived correctly** by web archives, such as the Internet Archive.

Who is it for?

Web professionals who need to check if their websites are archive ready.

Web archive engineers who need to evaluate target websites before harvesting and ingestion.

Students who need to learn more about digital archiving.

Testimonials

❝ *I teach a digital curation tools course, and one of the assignments involves web archiving. My students use ArchiveReady to determine potential problems with the sites they have chosen for their web archive collections. It's simple to use, and an exceptional tool!*
Alyce L. Scott, Lecturer, School of Information, San Jose State University

Service started in 2012/10/01, websites checked: 171257, software version: 4.1.

Why bother?

Because **not all websites are archive ready** and this may result in invalid web archives and ultimately in **information loss**.

How does it work?

ArchiveReady analyses your website (i.e. HTML, Images, CSS, JS, Sitemaps) and performs complex evaluations in order to calculate **Website Archivability** from a set of Archivability Facets: Accessibility, Cohesion, Metadata & Standards Compliance.

Do you need more information?

Banos V., Manolopoulos Y., Web Content Management Systems Archivability, ADBIS 2015. BIB.

Banos V., Manolopoulos Y.: A quantitative approach to evaluate Website Archivability using the CLEAR+ method, International Journal on Digital Libraries, 2015, Springer Link, BIB.

Banos V., Kim Y., Ross S., Manolopoulos Y.: CLEAR: a credible method to evaluate website archivability, iPRES 2013, PDF, BIB.

How much does it cost?

Its **completely FREE for personal use**. Large scale use of the ArchiveReady.com API as well as technical and scientific support is available on a fee. Please contact the author.

Checking website: http://ccaha.org

All messages

📊 Summary   **✔HTML and CSS 8**   **✔HTTP 4**   **✔Media 9**   **✔Sitemaps 3**

---

01. **⚠ Local images found: 4, remote images found: 1**

used in a website but hosted elsewhere may cause problems in web archiving because they may not be captured when the site is archived. What is more, the site is depended on 3rd party sites which may not be available in the future, causing incontrolable problems.

4 local :

http://ccaha.org/sites/default/files/styles/medium/public/2018-04/treating-artifacts.png?h=cc872d96&itok=tzys_NCs

http://ccaha.org/sites/default/files/styles/medium/public/2018-04/providing-education.png?h=cc872d96&itok=GziuBv-G

http://ccaha.org/themes/ccaha/logo.svg

http://ccaha.org/sites/default/files/styles/medium/public/2018-04/preserving-collectings.png?h=cc872d96&itok=qO2VPQgq

1 remote :

https://widgets.guidestar.org/TransparencySeal/7107365

Reference: http://en.wikipedia.org/wiki/Inline_linking

📊 Rating 80%   **📶 Significance: ★★**   **Cohesion**

02. **ℹ Checking 5 images with JHOVE Object Validation Environment**

Format validation is the process of determining the level of compliance of a digital object to the specification for its purported format, e.g.: I have an object purportedly of format F; is it?

Reference: http://jhove.sourceforge.net/

Standards Compliance

---

03. **☑ Image checked: http://ccaha.org/sites/default/files/styles/medium/public/2018-04/providing-education.png?h=cc872d96&itok=GziuBv-G**

Jhove (Rel. 1.28.0, 2023-07-15) Date: 2024-11-13 21:22:09 GMT+01:00 RepresentationInformation: /dev/shm/tmp9_lfcdxi ReportingModule: BYTESTREAM, Rel. 1.4 (2018-10-01) LastModified: 2024-11-13 21:22:09 GMT+01:00 Size: 41422 Format: bytestream Status: Well-Formed and valid MIMEtype: application/octet-stream

📊 Rating 100%   **📶 Significance: ★★**   **Standards Compliance**

---

04. **☑ Image checked: http://ccaha.org/sites/default/files/styles/medium/public/2018-04/treating-artifacts.png?**

TOOLS AND VENDORS

ARCHIVE-IT

COMMUNITY WEBS

INTERNET ARCHIVE

Community Webs Membership Application

Internet Archive accepts applications from public libraries and other community-focused cultural heritage organizations interested in participating in an international network engaged in local history preservation. For more information see the call for applications or contact communitywebs@archive.org.

Name *

First Name

Last Name

ARCHIVE WEBPAGE

Interactive browser-based web archiving from Webrecorder.

The ArchiveWeb.page browser extension and standalone application allows you to capture web archives interactively as you browse!
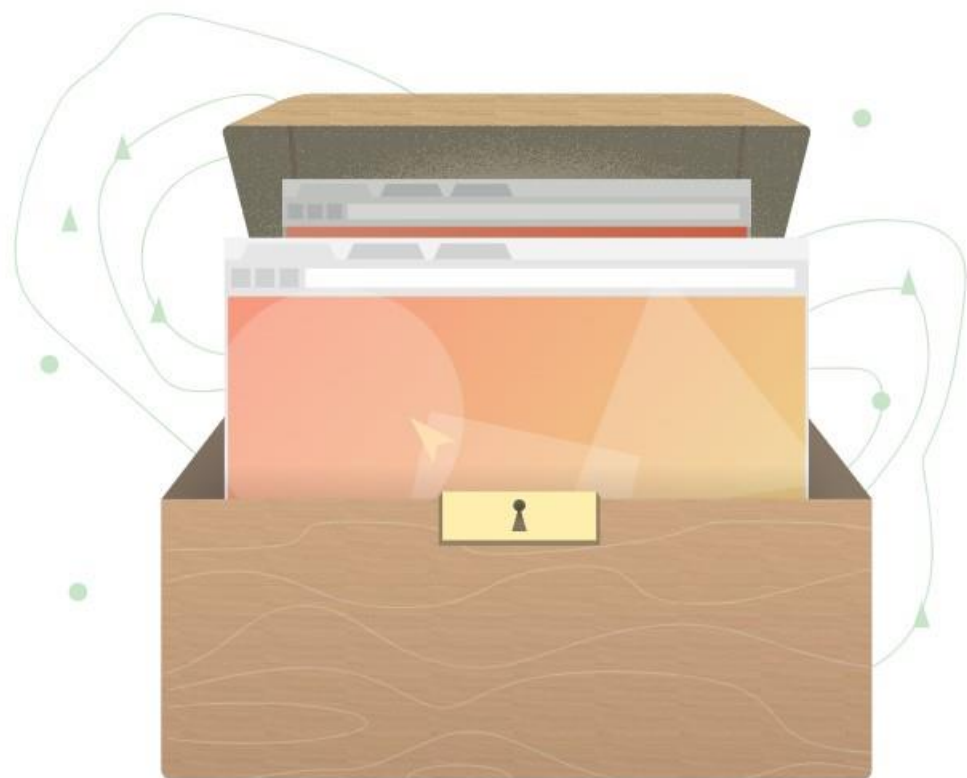
After archiving your webpages, your archives can be viewed using ReplayWeb.page — no extension required!

For those who need to crawl whole websites with automated tools, check out Browsertrix.

View the ArchiveWeb.page User Guide

Install Extension from Chrome Web Store

Download ArchiveWeb.page Desktop App - Latest version available for download!

Conifer

Collect and revisit web pages.

Conifer is a web archiving service that creates an interactive copy of any web page that you browse, including content revealed by your interactions such as playing video and audio, scrolling, clicking buttons, and so forth.

Create a Free Account

Existing Users Login

Free accounts with 5GB of storage. Get more and support this project by becoming a supporter.

Conifer is an online service based on Webrecorder software.

SOCIAL MEDIA ARCHIVING
(OH NO!)

Access and download your information on Instagram

🔗 **Copy link**

Android App Help   **iPhone App Help**   **Computer Help**   **Mobile Browser Help**   **More** ▾

You may not have access to this setting in Accounts Center at this time. If you don't see this setting under **Your Information and Permissions** in Accounts Center, try accessing this setting through your Instagram settings.

You're in control of your information on Instagram. You can view your account information or download a copy of your information on Instagram at any time.

Review your information in Accounts Center:

1. Click ☰ **More** in the bottom left, then click **Settings** ⚙.

2. Click **Accounts Center**, then click **Your information and permissions**.

3. Click **Access your information**.

You can also access this setting through your Instagram settings.

MORE ADVANCED TOOLS

- Web Curator Tool (National Library of New Zealand and National Library of the Netherlands)

- NetArchive Suite (Royal Danish Library, National Library of France, and National Library of Austria)

- Minet social media data scrapping

REFERENCES AND RESOURCES