Help:Using the Internet Archive

My name is Shawn Douglas, and I'm the curator and a senior editor here at LIMSwiki. One of the major tasks we undertake with the wiki is academic/historic research and writing of articles related to laboratory informatics, with the goal of providing useful information to those interested in the field. Traditionally, academic and historic research used to entail digging through physical stacks and archives in libraries and store rooms. Books, magazines, newspapers, journals, brochures, and other grey literature have long played an important role of not only learning more about specific topics but reconstructing fragments of history into a coherent whole.

The advent of the Internet and improved computing and storage technology, however, has brought with it new ways to create, publish, and archive. Books, film, music, news, and more have become staples of the Internet and other computer networks around the world, their popularity spurred by cheaper, more readily available digital publishing tools. Yet just as readily as new material is being uploaded and published to the Internet in large quantities, an alarming amount of digitally published material is either replaced or removed forever. This rapid and voluminous creation and destruction of mundane and creative cultural material is forcing researchers, historians, and data preservationists of all sorts to further examine what should be archived and how it should be done. One of many important tools to evolve from this examination is the Internet Archive [1].

What is the Internet Archive and why is it important?
The Internet Archive is a non-profit entity with the goal of building an Internet-based library. The non-profit describes why it's doing this as such:

Libraries exist to preserve society's cultural artifacts and to provide access to them. If libraries are to continue to foster education and scholarship in this era of digital technology, it's essential for them to extend those functions into the digital world.

Many early movies were recycled to recover the silver in the film. The Library of Alexandria — an ancient center of learning containing a copy of every book in the world — was eventually burned to the ground. Even now, at the turn of the 21st century, no comprehensive archives of television or radio programs exist.

But without cultural artifacts, civilization has no memory and no mechanism to learn from its successes and failures. And paradoxically, with the explosion of the Internet, we live in what [Applied Minds' Chief Technology Officer] Danny Hillis has referred to as our "digital dark age."

The Internet Archive is working to prevent the Internet — a new medium with major historical significance — and other "born-digital" materials from disappearing into the past. Collaborating with institutions including the Library of Congress and the Smithsonian, we are working to preserve a record for generations to come.
External audio


The Internet Archive's digital library includes more than seven million texts, two million audio recordings, and nearly two million videos that have fallen or intentionally been published to the public domain. The non-profit also has another important tool: the Wayback Machine [3]. This tool functions as "a three-dimensional index that allows browsing of web documents over multiple time periods," and it contains nearly two petabytes of archived web data. The Wayback Machine gives researchers the ability to see past iterations of a website as long as the web address is known. For example, the now defunct X-Files website can still be viewed in its various forms dating back to 1996. (It also includes a bit of history for X-Files buffs: the original owner of the domain allegedly received legal threats from Fox in September 1997 concerning the domain.) And that previous fact in the parentheses? I was able to add a citation to that statement all thanks to the Internet Archive!

So why is the Internet Archive important? Well, hopefully the previous example of The X-Files mildly illustrates the importance of the service. From someone who's researching the history of the television series to an editor working on The X-Files Wikipedia entry [4], having access to content that was originally published at the associated web domain — via the Wayback Machine — is particularly useful, not only for finding facts but also citing them. LIMSwiki editors also make good use of the Wayback Machine in their research of laboratory informatics vendors and open-source software projects: many companies and projects either alter their web content or disappear from the Internet completely, with only the Internet Archive to provide clues. Finally, cultural anthropologists, data preservationists, and historians aren't the only ones discovering cultural records on the site; people from all walks of life are tapping into both "born-digital" and digitized physical materials, some of which date back multiple centuries. From a 1929 recording of Louis Armstrong & His Orchestra's "Ain't Misbehavin'" [5] to an English dictionary [6] published in 1720, the Internet Archive gives people from all walks of life a chance to revisit a cultural past old and recent.

Activities

Let's learn a bit more about the Internet Archive and its offerings, performing a few activities in the process.

Internet Archive as an Internet library

1. Learning about the Internet Archive's approach to data preservation

Open the following YouTube video in a new browser tab and watch it: Internet Archive by Deepspeed Media [7]

Starting at 2:48, Brewster Kahle, founder of the Internet Archive, explains several problems with data preservation that the non-profit has had to deal with.

1a. Explain at least two of the considerations the Internet Archive have made in regards to preserving books, videos, audio, etc.

1b. At 6:20 Robert Miller, Global Director of Books, explains that there's more than just a digital preservation effort. What is the Internet Archive doing with physical books? Do you think it's a worthy effort? Why?
Citing files from the Internet Archive

As you have learned, the Internet Archive is quite serious about its efforts to bring the world digitally preserved physical and "digital-born" content. Texts in particular are useful to researchers, and the collection of over seven million texts found at the Archive make for a wonderful trove of research material. But how do we cite that material?

On Wikipedia, this wiki, or other wikis with citation tools

On this and other wikis, citations are required. Citations are placed via citation templates. Here are a few unpopulated citation templates, for example:

- \((\text{cite web} | \text{url}= | \text{format}= | \text{title}= | \text{work}= | \text{author}= | \text{publisher}= | \text{date}= | \text{accessdate}=))\)
- \((\text{cite book} | \text{url}= | \text{chapter}= | \text{title}= | \text{author}= | \text{pages}= | \text{publisher}= | \text{year}= | \text{edition}= | \text{volume}= | \text{isbn}= | \text{accessdate}=))\)
- \((\text{cite journal} | \text{url}= | \text{format}= | \text{journal}= | \text{chapter}= | \text{title}= | \text{author}= | \text{year}= | \text{volume}= | \text{issue}= | \text{pages}= | \text{pmid}= | \text{doi}= | \text{accessdate}=))\)
- \((\text{cite news} | \text{url}= | \text{format}= | \text{title}= | \text{author}= | \text{agency}= | \text{publisher}= | \text{newspaper}= | \text{pages}= | \text{location}= | \text{date}= | \text{accessdate}=))\)

This guide isn't dedicated to showing you how to create citations; consult the advanced training section of the MediaWiki training guide to learn more. That said, citing a digitally archived book is straightforward. Using the \{\text{cite book}\} template, you'll be required to make a few modifications. Let's use that 1720 dictionary as an example, focusing on the definition of "algorithm":

\[
\text{Dict1720} = \text{cite book} | \text{url}= https://archive.org/details/The_New_World_of_English_Words_Or_A_General_Dictionary | \text{chapter}=Algorithm | \text{title}=The New World of Words: Or, Universal English Dictionary | \text{author}=Phillips, Edward | \text{publisher}=Internet Archive | \text{others}=Taylor, I. | \text{date}=1720 | \text{edition}=7th | \text{accessdate}=25 November 2014)
\]

In this case, the default \{\text{cite book}\} template doesn't include room for the document location or the name of the archive. In this example, I've put the publisher I. Taylor in the "others" (other contributors) field and put the archive name in the publisher field so it would appear in a more realistic spot. It would appear as such in a references section of the wiki page:

2. Making an archived book citation in a wiki

You've found the story "The Spirit of the Willow Tree" in a book called Ancient Tales and Folklore in Japan. You need to cite this story, located in chapter two, pages 12–18. The source is found here: https://archive.org/details/ancienttalesandf00gordrich

Create a full wiki citation for this archived book using the guidance above.

In a research paper

MLA: This comes from the seventh edition of Rules for Writers by Diana Hacker and Nancy Sommers:

Digital archives are online collections of documents on records — books, letters, photographs, data — that have been converted to digital form. Cite publication information for the original document, if it is available... Then give the location of the document, if any, neither italicized nor in quotation marks; the name of the archive, italicized; the medium ("Web"); and your date of access.

One of the examples Hacker and Sommers uses:

Help: Using the Internet Archive

Here's an example using the Internet Archive's digitized book *Ancient Tales and Folklore in Japan* from activity two:


Using the Internet Archive's digitized book *Ancient Tales and Folklore in Japan* from activity two:


Open the following YouTube video in a new browser tab and watch it: Internet Archive’s Wayback Machine. User Willie D. explained how to use the Wayback Machine and stated one reason for using it as it being “fun.”

1a. Did you find the Wayback Machine interface easy-to-use or difficult? Explain.

1b. What other uses can you imagine for using the Wayback Machine other than for fun? State two real-world problems or activities that the Machine could be applied to and explain how it would resolve each problem or benefit each activity.

Imagine you're researching the history of Keane International, an IT services and solutions company. In your research you discover Keane was officially acquired by NTT DATA Corporation on January 3, 2011, and you find its former web domain was http://www.keane.com/.

5a. Using the Wayback Machine and the Keane domain name, find the following pieces of information:

* the first available archive date for the domain
Help: Using the Internet Archive

* the first-quarter 1996 revenues as reported to investors
* the name of the founder of the company
* the year the company acquired GE Consulting Services

Citing web pages from the Wayback Machine

On Wikipedia, this wiki, or other wikis with citation tools

Using the information from citing files on a wiki, citing an archived webpage is also rather straightforward. Using the {{cite web}} template, you'll be required to add a few additional parameters: "archiveurl" and "archivedate". Let's use the November 25, 2005 version of the Keane website as an example:


In this case we added the "archiveurl", which is the URL for that archived page found in the browser address window, and the "archivedate", the archive date we selected from the interface. Note that if you forget the "archivedate" parameter, the system will show an error message in the citations stating "Error: If you specify |archiveurl=, you must also specify |archivedate=".

6. Making a citation in a wiki

You've found the 1994 shareholder equity number for IBM using the Wayback Machine: $23,413,000,000. The source is this page: https://web.archive.org/web/19961026210557/http://www.ibm.com/IBM/ibmar94/finance4.2.html

Create a full wiki citation for this archived webpage using the guidance above.

In a research paper

MLA: This commentary on MLA formatting comes directly from the Internet Archive website:

This question is a newer one. We asked MLA to help us with how to cite an archived URL in correct format. They did say that there is no established format for resources like the Wayback Machine, but it's best to err on the side of more information. You should cite the webpage as you would normally, and then give the Wayback Machine information. They provided the following example: McDonald, R. C. "Basic Canary Care." _Robirda Online_. 12 Sept. 2004. 18 Dec. 2006 [http://www.robirda.com/cancare.html]. _Robirda_.

Internet Archive_. [http://web.archive.org/web/20041009202820/http://www.robirda.com/cancare.html]. They added that if the date that the information was updated is missing, one can use the closest date in the Wayback Machine. Then comes the date when the page is retrieved and the original URL. Neither URL should be underlined in the bibliography itself. Thanks MLA!

This information may be a bit outdated since as of 2012 the MLA Handbook states that web addresses are not required. They do have a particular format, however, if your professor requires a URL: use < and >. That said, I'd probably format the above quoted material as:


Note the formatting: author, last name first; title of the website; update date; access date; original URL; source of archived URL; archived URL. For more on MLA formatting, see the Purdue OWL MLA Style Guide [8].
APA: As for making the same citation in APA style, how to do it is less clear and authoritative. I’ve found at least one professor who states “just cite the web page where you found your information.” I would tend to agree, simply using the archive URL rather than the original:


Note the formatting: author, last name first; update date; title of the website; ”Retrieved from” archived URL. For more on APA formatting, see the Purdue OWL APA Style Guide [9].

7. Creating MLA and APA citations for an archived webpage

Use one of the Wayback webpages you encountered for Keane.com from activity two. State the fact you are citing (one of the bullet point items) and include both an MLA and APA citation of that webpage.

Handout

Citing from a Digital Archive like the Internet Archive: Cheat Sheet [13] (downloadable PDF)

Citing from a Digital Archive like the Internet Archive: Cheat Sheet [14] (online PNG file)

Associated help pages

• Help documentation for using MediaWiki

External links

• Internet Archive [1]
• Wayback Machine [3]

References

[7] https://www.youtube.com/watch?v=ec_-fgy3EGY#t=168
[8] https://owl.english.purdue.edu/owl/resource/747/01/
[9] https://owl.english.purdue.edu/owl/resource/560/1/
[14] http://i.imgur.com/zFMIxIg.png
Article Sources and Contributors


Image Sources, Licenses and Contributors


License

a Creative Commons Attribution-Share Alike 4.0 International License
LIMSWiki:Copyright
https://creativecommons.org/licenses/by-sa/4.0