

Researching the Internet Archive

Thomas MacEntee, of High-Definition Genealogy
<http://hidefgen.com> hidefgen@gmail.com

Even if you've been using more and more online resources for your genealogy research, you may not have heard of the Internet Archive, a hidden gem filled with solid resources for every genealogist.

What is the Internet Archive?

The Internet Archive is a non-profit, Web-based, digital library offering storage of, and access to, digitized materials. These materials include archived websites, books, videos, images, audio files and more. You can find the Internet Archive at <http://www.archive.org>. In addition, the Internet Archive offers The Wayback Machine which allows you to see what a website looked like 2, 5 or even 10 years ago!

The Internet Archive is important to genealogy research for several reasons:

- Most of the items are in the public domain, meaning they can be freely used and copied.
- Many genealogy books and even microfilm images of census sheets and more are available.
- These resources are available to the public for free.

The Wayback Machine

The **Wayback Machine** (<http://archive.org/web/web.php>) is an archive of how website pages appeared at certain time periods. When the Internet got started, no one at first thought it would be important or of value to have such information. However, having a "snapshot" of what a website looked like is valuable.

1. Enter a **website address** or **URL** in the search field and click **Take Me Back**.
2. A timeline appears for that website with dates on which a "snapshot" was taken. Click on a **date** to view the snapshot.

The Wayback Machine can be used to find those "broken links" and pages that no longer seem to exist.

Books and Texts

The Texts collection contains almost 4 million books in digital form and of all the collections at Internet Archive, has the most relevant items for genealogists and family historians.

1. On the **Menu Bar**, select **Texts**. The **Text Archive** page appears.
2. In the **Search** field enter your keyword(s) and click **Go**.

Audio and Video

There are over 1 million video items and over 1.5 million audio recordings at Internet Archive.

1. On the **Menu Bar**, select **Video** or **Audio**. The **Video** or **Audio Archive** page appears.
2. In the **Search** field enter your keyword(s) and click **Go**.

Search Strategies

Take advantage of the Advanced Search screen when possible. The form can be a bit daunting and is very complex but it can help pinpoint materials especially for those common surnames.

Downloading Content

There is an extensive FAQ (“frequently asked questions”) section about searching and downloading books on the Internet Archive. Here are some tips :

- Once you’ve located a text, on the left sidebar is a grouping of links labeled **View** the book.
- The most common text formats are included such as PDF, EPUB or for e-book readers like Amazon’s Kindle.
- Read Online** displays a viewer that allows you to “flip” pages, zoom in on content and even search within the book.
- The viewer also allows you to share the book via social media with others and even “listen” to an audio reading of the book.
- A PDF format is available for most texts but it will take a long time to download depending upon your Internet access speed and the size of the book.
- Right-click over the PDF link, select **Save Link As**, and save the PDF to your computer. Then open the PDF for viewing.
- The Full Text format will download a small .txt formatted file but with lots of extraneous coding.
- The Full Text format is the quickest to download and easiest to search but lacks structure, pagination and other conveniences.

Uploading Content

Many Internet Archive users don’t realize that you can create a user account on the site and then upload content to add to the Internet Archive library. Click the Upload button in the upper right corner – you will need to have an Internet Archive account.

Hidden Gems of the Internet Archive Collection

Did you know . . .

- That several sets of US Federal Census forms and other censuses are available in the Text Collection of the Internet Archive?
- That Internet Archives has a special genealogy section? See <http://www.archive.org/details/genealogy>

The Genealogy Collection includes:

- Items from the Allen County Public Library, the National Library of Scotland and the Boston Public Library.
- Almost 2,000 compiled family genealogies.
- Over 3,000 birth registers; and more!

Internet Archives Tips and Tricks

Here are some tips on the best ways to use the Internet Archive for genealogy research:

- Find dead web pages.** When you encounter a “page not found” or “404” error on a website you’ve visited in the past, copy the page address and paste it into The Wayback Machine search engine. Then use the timeline to locate the last snapshot of that page.
- Not everything is in The Wayback Machine.** While Internet Archive does a good job of taking periodic snapshots of web pages, it only does so for important or top level pages. You may find that sub-pages are not included.
- Search broad then go narrow.** Even when working with common surnames or place names, do a broad search and then review the results. Next add unique words or qualifiers such as “genealogy” or dates.
- Not all content is searchable.** While many texts have been “OCRd” meaning the image has been converted to text, not all books or documents have been processed. You may have to browse through the pages of a text.
- Search by collection.** When reviewing search results, look in the sidebar on the right and use one of the Collection links displayed. Also keep scrolling down to use the Related Collections links.
- Use Google to search the Internet Archive:** Preface your search with “site:archive.org” in Google and only those results found at Internet Archive will be displayed.

Links

- ❑ A User's Guide to the Internet Archive
<http://bit.ly/hdgiauserguide>
- ❑ Brewster Kahle's Internet Archive
<http://bit.ly/iabkinterview>
- ❑ **Internet Archive**
<http://archive.org/index.php>
- ❑ **Internet Archive – Blogs**
<http://blog.archive.org>
- ❑ **Internet Archive – FAQs**
<http://archive.org/about/faqs.php>
- ❑ **Internet Archive – Audio**
<http://archive.org/details/audio>
- ❑ **Internet Archive – Texts**
<http://archive.org/details/texts>
- ❑ **Internet Archive – Video**
<http://archive.org/details/movies>