Curating the Internet: The Collaborative Future of Library Collections

Fiesole Collection Development Retreat
Lille, April 20, 2017
Michael Levine-Clark
University of Denver
We concentrate our efforts on a very small portion of a huge potential collection.
We concentrate our efforts on a very small portion of a huge potential collection.
Expanding the Collection

We concentrate our efforts on a very small portion of a huge potential collection
+300 million files preserved. Adding more content every day.

What’s in the Archive? »
Low Stewardship

High Stewardship

In Many Collections

In Few Collections

Low-Low
Freely-accessible web resources
Open source software
Newsgroup archives

High-High
Special Collections
Rare books
Local/Historical Newspapers
Local History Materials
Archives & Manuscripts
Theses & dissertations

High-Low
Research & Learning Materials
Institutional records
ePrints/tech reports
Learning objects
Courseware
E-portfolios
Research data
Prospectus
Insititutional website

COLLECTIONS GRID
(Lorcan Dempsey and Eric Childress, OCLC Research)
Internet Archive is a non-profit library of millions of free books, movies, software, music, websites, and more.

Search the history of over 286 billion web pages on the Internet.
About the Internet Archive

The Internet Archive is a 501(c)(3) non-profit library. Founded in 1996, our mission is to provide Universal Access to All Knowledge. We collect published works and make them available in digital formats. We are building a public library that can serve anyone in the world with access to the Internet.

We began in 1996 by archiving the Internet itself, a medium that was just beginning to grow in use. Like newspapers, the content published on the web was ephemeral - but unlike newspapers, no one was saving it. Today we have 20+ years of web history accessible through the Wayback Machine and we work with 450+ library and other partners through our Archive-It program to identify important web pages.

As our web archive grew, so did our commitment to providing digital versions of other published works. Today our archive contains:

- 279 billion web pages
- 11 million books and texts
- 4 million audio recordings (including 160,000 live concerts)
- 3 million videos (including 1 million Television News programs)
- 1 million images
- 100,000 software programs

Anyone with a free account can upload media to the Internet Archive. We work with thousands of partners globally to save copies of their work into special collections.
Not in Archive

The page you requested has not been archived in Archive-It.

This could be for a number of reasons.

Most likely the page you are requesting was outside of the crawler's scope. Try another request or click to see other pages from news.du.edu.

It is also possible that this site is currently being crawled and the archived pages are not yet available in the Wayback Machine. It can take up to 24 hours after a crawl has been completed for the site to appear in the Wayback Machine. Please try again after the allotted time.

You can also try searching for news.du.edu/new-partnership-with-denver-health-will-help-expectant-mothers/ on the live web, in the global Wayback Service, or in the General Archive at the Internet Archive at archive.org
Hrm.

Wayback Machine doesn’t have that page archived. Want to search for all archived pages under https://www.greatamericanbeerfestival.com/?

This page is available on the web!
Help make the Wayback Machine more complete!
Save this URL in the Wayback Machine
How Tweet It Is!: Library Acquires Entire Twitter Archive

April 14, 2010 by Matt Raymond
Activists Rush to Save Government Science Data — If They Can Find It

By AMY HARMON  MARCH 6, 2017
Building refuge for federal climate & environmental data

DataRefuge helps to build refuge for federal data and supports climate and environmental research and advocacy. We are committed to fact-based arguments. DataRefuge preserves the facts we need at a time of ongoing climate change.

This site is one part of the project. The vast majority of the government information gathered through this project is available from the Internet Archive through the End of Term project. This data catalog is a place to store data that is difficult or impossible to harvest through web crawlers. Beginning to add difficult data to this websites is one of the tasks of the DataRescuePhilly event on January 13th and 14th.
Dépôt légal des sites web

Types de documents

Le dépôt légal concerne tous les types de publications diffusées sur l'internet : sites institutionnels ou personnels, revues d'accès gratuit ou payant, blogs, sites commerciaux, plateformes de vidéos ou de livres numériques...

Les contenus édités ou les éditeurs doivent avoir un lien avec le territoire français. Il s’agit avant tout des sites web enregistrés en .fr ainsi que ceux des départements et territoires d’outre-mer. D’autres sites web en .com, .org, .net, etc. édité par des personnes physiques ou morales domiciliées en France font également partie du périmètre. Enfin, les publications produites sur le territoire national peuvent être concernées même si elles sont diffusées par une société étrangère.

Afin d’identifier ces sites web, la BnF a passé des accords avec des organismes en charge de la gestion de noms de domaines, comme l'Association française pour le nommage Internet en coopération (AFNIC), la société de conseil en systèmes et logiciels informatiques (OVH) et l’Office des postes et télécommunications de Nouvelle-Caledonie (OPT-NC). Chaque année, plusieurs millions de sites sont ainsi capturés, jusqu’à une profondeur déterminée à l’avance (de l’ordre de quelques milliers de fichiers par site).

http://www.bnf.fr/fr/professionnels/depot_legal/a.dl_sites_web_mod.html

Web Archiving at the Library of Congress

In 2000, the Library of Congress established a pilot project to collect and preserve websites. A multidisciplinary team of Library staff studied methods to evaluate, select, collect, catalog, provide access to, and preserve these materials for future generations of researchers.

Since then, the Library has developed thematic web and event-based archives on such topics as the United States National Elections, the Iraq War, and the events of September 11. Designated staff in Library Services and the Law Library select content for the archives.

A Web Archiving Team, currently located in the Library’s Technology Policy office, supports the goal of managing and sustaining at-risk digital content. The team is charged with building a Library-wide understanding and technical infrastructure for capturing web content. The team, in collaboration with a variety of Library staff and national and international partners, is identifying policy issues, establishing best practices and building tools to collect and preserve web content.

https://www.loc.gov/webarchiving/about.html
Das müssen Sie wissen!

Das Leben in Deutschland und den Terroranschlägen, die sich in der Folge von Diskussionen die Frage nach Gewalt und Frieden, die Täler und ihre Kämpfer: Ein speziell umfassenden Hintergründe, Analysen und Erfahrungsberichten >>

Der 11. September 2001 und die Folgen

Vier Französische Schulen
(Zusammenstellung)

AUCTION

Das aktuelle Gebot: 50 Euro

Ende der Auktion: 12.9.2002, 13 Uhr

Der Cartoonist Art Spiegelman hat die Terroranschläge des 11. September in der Zeitschrift "Der 11. September 2001 und die Folgen" und "AUCTION".
A-Z Databases: Newspapers
Find the best library databases for your research.

104 Databases found for Newspapers

Access World News
Provides extensive news coverage at any level—local, state, regional, national and international. Features the vast majority of U.S. newspapers by circulation, along with almost one thousand hard-to-find local and regional titles, and offers searchable news video clips and over two thousand international news sources from scores of countries on six continents, translated into English when written in other languages.

African American Newspapers, 1827-1998
Provides access to U.S. newspapers from more than 35 states chronicling a century and a half of the African American experience.

African Newspapers (World Newspaper Archive)

America's Historical Newspapers
Thousands of titles chronicle centuries of American history, culture and daily life. Includes titles from all 50 present states. This database provides access to digitized versions of primary source material.

American Business: Agricultural Newspapers
Historical newspapers covering agriculture, agricultural technology, and economics of farming published between 1788 through 1894.
Karen Women Organisation

Archived content not available outside of Library of Congress premises: this archived web site content displays only at the Library of Congress.
ALLY SLOPER TACKLES THE EASTERN QUESTION. Sixpence.

ALLY SLOPER'S

COmIC

1878

CALENDaR

ONE PENNY

Judy" Office, 78 Fleet Street, London, E.C.

16 Miniature Souvenir Views
AMERICA'S EXPOSITION
SAN DIEGO, CALIFORNIA

...
How do libraries preserve the primary sources on the Internet?

• A constantly moving target

• Too much to ever preserve in depth

• Collaboration
  • To preserve some resources at the broad/surface level
    • Are the Internet Archive and International Internet Preservation Consortium doing this well enough?
  • To preserve some resources in depth
    • Is this the role for libraries?

• Legal and ethical issues
Marijuana in Colorado

Recreational and Medical Marijuana Stores in Denver

Denver is the largest city in Colorado, as well as the state capital. In Denver, you will find the highest concentration of recreational marijuana stores in Colorado, with a large number of select stores selling recreational and medical marijuana, 11 selling recreational only, and 21 medical only. Marijuana stores in Denver are required to close by 7pm. You may want to explore the area on a bike, a rental bike system offered throughout the city.
Facts & Figures

As of 07/2016, the WEST Archives Include:

- 13817 Backfiles
- 20400 Titles
- 504435 Volumes
Area-wide transition to open access is possible

A new study calculates a redeployment of funds in Open Access

April 27, 2015

The Max Planck Digital Library has put forward a study on the transformation of the subscription-driven system for scientific publications to an Open Access model. For the first time, quantitative parameters are presented showing that the liberation of scholarly literature is possible at no extra costs.

According to market analyses, annual turnovers of academic publishers amount to approximately EUR 7.6 billion. This money comes predominantly from publicly funded scientific libraries as they purchase subscriptions or licenses in order to provide access to scientific journals for their customers. Since more than a decade the Open Access movement, in which the Max Planck Society plays a major role, has been demanding free and immediate access to the results of academic research on the internet.
To preserve the Internet, libraries should... 

- Identify locally-important websites to preserve in depth
- Work collaboratively
- Expend a bit less money and effort on “traditional” resources