



Archive-it



- What
 - Web based service that allows partners to create, manage and search their web archives through an easy to use web interface
 - Functions include: Harvesting, cataloging, managing of collections, search and monitor
 - Designed for users without much technical infrastructure or support to run such tools internally
 - Utilizes all open source tools:
Nutchwax/Heritrix/Wayback and ARC file format



Why

- Requests from multiple institutions to harvest material not currently in the wayback
- Wanted more control and access capability
- Wanted to expand crawling/archiving capability to a much broader range of institutions cost effectively



Project

- Launched first concept in March 2005
- Finished initial set of design requirements May 2005
- Signed on pilot partners June 2005
- Launched pilot and beta application September 2005
- End pilot November 18th 2005
- Launch first production release in January 2006



Pilot Partners

- Center for Research Libraries
- Research Libraries Group (U of Toronto, U of Indiana, Haverford and Swarthmore Colleges, IISH)
- University of Texas
- Library of Virginia
- State Archives South Dakota
- State Archives North Carolina
- State Archives Alabama
- Minnesota Historical Society
- Institut d'Etude Politique de Grenoble



Funding

- Funded internally for pilot no cost to partners
- Subscription based on full release
 - \$10,000 per year, up to 10 million documents per account (roughly 50-100 sites, once per month)

Archive-it

Grab File Edit Capture Window Help

Archive-It

http://www.archive-it.org/index.html

Getting Started Latest Headlines Archive-It Archive-It: Member L... BNF Crawler Metrics WERA-0.2.3-200510... Archive-It! home.faces Survey Summary Archive-It JIRA Minerva Web Harvest...

ARCHIVING THE INTERNET FOR FUTURE GENERATIONS
COLLECT IT, MANAGE IT, SEARCH IT...ARCHIVE-IT

Member Log In

ARCHIVE-IT Enter Search Terms Here! Search All Collections Go

Learn More About Us Pilot Partners FAQ Contact Us

Featured Collections

Katrina Hurricane
<http://www.archive.org>
The Internet Archive, Library of Congress and many individual contributors worked together to put together a comprehensive list of websites to create a historical record of the devastation and massive relief effort, due to Hurricane Katrina. This collection is a sampling of the larger web archive available at Internet Archive.

Virginia's Political Landscape, Fall 2005
http://www.lva.lib.va.us/whatwehave/archive_web.htm
The Library of Virginia's "Virginia's Political Landscape, Fall 2005" archiving project documents political activity in the Commonwealth during the last 4 months of the administration of Governor Mark R. Warner. This includes the Web sites of those running in the state-wide elections of the fall of 2005, as well as related political sites, in an effort to record the often ephemeral activities of political candidates.

Latin American Government Documents Archive
<http://lanic.utexas.edu/project/archives/lagda/>
The Latin American Government Documents Archive (LAGDA) seeks to preserve and facilitate access to a wide range of ministerial and presidential documents from 21 Latin American and Caribbean countries. The Archive contains copies of the Web sites of over 100 government ministries and presidencies, captured on multiple dates during a two month period in the Fall of 2005.



Other Features planned

- Search across multiple collections
- Improved search by date range/mime type/date order
- Collection collaboration among partners
- Individual user account and permissions for each account
- Metadata search and filtering of text search
- Ability to set crawl parameters by collection
- Reporting by seed and QA interface
- Integration of discovery process