

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



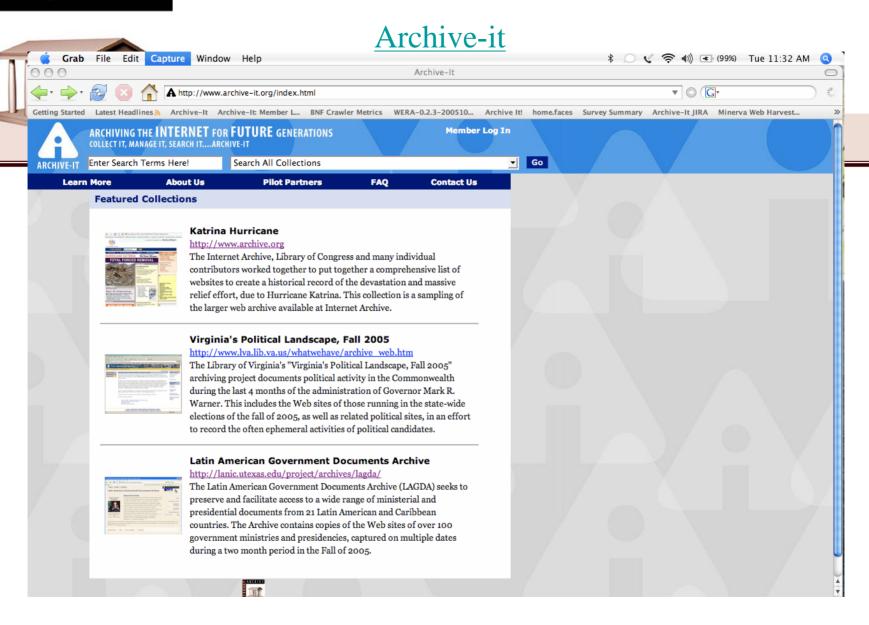
Pilot Partners

- 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)





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