



State Library of Massachusetts
Electronic Documents Archive

Electronic Documents Stored in DSpace

- Open source Institutional Repository software developed in 2002 by MIT and HP to store theses and dissertations
 - Robust but bare-bones system
 - Written and customizable in Java
 - Provides full-text keyword searching of documents
 - Several hundred institutions (mostly universities) use DSpace throughout the world (partial list at www.dspace.org)



History of the Project

History of the Project

- In 2003 the library received initial funding to set up system to permanently store electronic documents
 - Difficulty locating documents on individual agency websites
 - No permanence – sometimes documents disappear
 - State law requires agencies and legislative offices to send State Library copies of any publications they produce

History of the Project

- Received state IT capital funds in 2003 to:
 - Configure a webcrawler that would locate and download .pdf and .doc files from agency sites
 - Create a database that manages the status of these downloaded files
 - Purchase server to permanently store state documents

History of the Project

This initial funding had also included software to enable keyword searching within documents.

This part of the project was cut due to funding constraints.

History of the Project

- Problems:
 - Webcrawler worked too well – difficult to manage the many, many documents that were retrieved
 - Software to manage items was cumbersome to use
 - Backlog in cataloging
- Took alternative approach:
 - As electronic items were discovered, they were cataloged with links to agency websites

History of the Project

- Second Phase of Funding – 2005 and 2006
 - DSpace chosen as electronic repository
 - State had strong preference for Open Source/Open Standards solutions
 - Worked with IT consultant to select and configure this product
 - Open source does not mean free – need IT expertise (JAVA) to configure
 - DSpace comes bundled with Lucene, which provides keyword indexing of documents

History of the Project

- Second Phase of Funding – 2005 and 2006
 - Scanning of Massachusetts Session Laws (Acts and Resolves) for 1960 to 1996
 - Approximately 50,000 pages
 - Each act is a separate file and fully keyword searchable
 - Used heavily by legislative staff, lawyers and town officials among others

Project Today

- Since 2007, we have been systematically adding “Digitally Born” content to dSpace
 - Ran report of OPAC records containing links to electronic documents
 - Downloaded these documents so library has permanent copy
 - Currently adding these documents to dSpace
 - Actively encouraging agencies to notify us about new digitally born reports (can send link or copy of document to email account)
 - Approximately 1000 items added to date

Project Today

- Staff is identifying and starting to add other scanned documents
 - The State Library (and several other area institutions) are actively scanning older state documents
 - As a member of the Boston Library Consortium, the library is scanning materials at the Internet Archive/Open Content Alliance's Scanning Center established at the Boston Public Library

Project Today

- Staff is identifying and starting to add other scanned documents
 - The library has scanned two large series
 - Legislative Biographical Directories (known as the Bird Books) running from about 1890 to the present
 - Annual Reports of the Attorney General (runs from the 1840's to the present)

Project Today

- Identifying other scanned documents
 - Collaborating with UMass Boston, UMass Amherst and Boston Public Library – sharing knowledge and plans

Project Today

- Collaboration in scanning documents
 - The Joseph P. Healey at University of Massachusetts Boston Library is sponsoring the digitization of our older Acts and Resolves (1600's to 1959)

Project Today

- Identifying other scanned documents
 - Other area institutions have scanned several series as well:
 - UMass Amherst – Yearly report on Vital Statistics, Election Statistics, Fruit Notes, and Annual Reports of Northampton State Hospital
 - Boston Public Library – Soldiers and Sailors of the Revolutionary War
 - Boston University – some years of the Department of Public Health

Project Today

- Identifying other scanned documents
 - The State Library has created a webpage with links to the major series scanned:
www.mass.gov/lib/collections/dc/StateDocumentsOnline.html

Project Today

- Identifying other scanned documents
 - More material from throughout the country is being scanned and added constantly to the Internet Archive site: www.archive.org



Future Plans

Future Plans

- Download archival copies of scanned documents and make available on dSpace (adds keyword search capability)
- Migrate and upgrade dSpace with state IT Department's help

Future Plans

- In the longer term, evaluate other digital asset management systems to see what best meets our needs
- Pull in our other smaller-scale digital projects so our digital assets are in one place

Contact Information

Alix Quan

Assistant Director/Head of Reference

State Library of Massachusetts

24 Beacon Street

Boston, MA 02133

617-727-2403

alix.quan@state.ma.us