DLXS at Cornell University

- Digital Collections at the Cornell University Library

- Migrating the Making of America Collection to DLXS v.13

- Local modifications

George Kozak
Cornell University Library IT
Cornell University
Ithaca, NY
gsk5@cornell.edu

Digital Collections @ CUL

- We have multiple digital collections at Cornell:

4 Greenstone collections (mostly newspapers such as the *Cornell Daily Sun*

34 Image Collections (using Luna Insight)

26 Text Collections using DLXS

Vivo – Virtual Science Library

eCommons@Cornell (DSpace)

arXiv - e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Digital Registry: http://rdc.library.cornell.edu

DLXS Text Collections @ CUL

- 26 DLXS Text Collections at Cornell
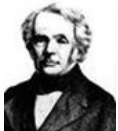- **Listed below from most hits to least:**

Making of America (MOA)

Southeast Asia Visions

Home Economics Archive (Hearth)

Core Historical Literature of Agriculture (CHLA)

Historic Math Book Collection

Text Collections @ Cornell

Kinematic Models for Design e-books

New York State Historical Literature

Historical Monographs

Witchcraft Collection

Samuel J. May Anti-Slavery Pamphlets

Cornell Modern Indonesia Collection

Text Collections @ CUL

Ezra Cornell Materials

Cornell University Digital Archives

Huntington Free Library of Native American Literature

Race, Ethnicity and Religion

Liberian Law

Study of Greek Revival Architecture

Text Collections @ CUL

Stevens Family Papers

Southeast Asia Program Data Series

Hive and the Honeybee Collection

History of the Left in Latin America

Cornell East Asia Series Online

A History of Social Welfare

Text Collections @ CUL

Bolivian Digital Pamphlets

Regmi Research Series

International Women's Periodicals

- *Our newest collection:*

  The Records of the Albany Commissioner of Indian Affairs, 1678-1755: An Integrated Digital Database

Cornell's Making of America

- 19th Century Literature from the antebellum period through reconstruction

- 22 journals (955 volumes)

- 114 monographs

- 165 volumes of Official Civil War Documents

Cornell's Making of America

- Collection was originally made available online in 1996.

- Originally we used locally developed code that used the Dienst Protocol (also used for Historic Monographs, Historic Math Books and NY State Historic Literature)

- In 1999, we switched to using the Michigan Middleware (pre-DLXS) for the Making of America Collection.

- 1,470 SGML files and 950,000 images

Cornell's Making of America

- Original interface

Cornell's Making of America

- Original Search Page

Cornell's Making of America

- Original Browse Page

Browse Cornell University's MoA Journal Collection

- The American Missionary (1878 - 1901)
- The American Whig Review (1845 - 1852)
- The Atlantic Monthly (1857 - 1901)
- The Bay State Monthly (1884 - 1886)
- The Century (1881 - 1899)
- The Continental Monthly (1862 - 1864)
- The Galaxy (1866 - 1878)
- Harper's New Monthly Magazine (1850 - 1899)
- The International Monthly Magazine (1850 - 1852)
- The Living Age (1844 - 1900)
- Manufacturer and Builder (1869 - 1894)
- The New England Magazine (1886 - 1900)
- The New-England Magazine (1831 - 1835)
- New Englander (1843 - 1892)
- The North American Review (1815 - 1900)
- The Old Guard (1863 - 1867)
- Punchinello (1870)
- Putnam's Monthly (1853 - 1870)
- Scientific American (1846 - 1869)
- Scribner's Magazine (1887 - 1896)
- Scribner's Monthly (1870 - 1881)
- The United States Democratic Review (1837 - 1859)

Browse Cornell University's MoA Multivolume Monographs

- Official Records of the Union and Confederate Navies in the War of the Rebellion (1894 - 1922)
- The War of the Rebellion: a Compilation of the Official Records of the Union and Confederate Armies (1880 - 1901)

Browse a complete Bibliography of Books and Articles in MoA, organized by author and title

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | Non-alphabetic

Cornell's Making of America

- Needed to fix known problems:
  - Duplicated and misplaced images
  - Missing images
  - Missing pages and volumes
  - Incorrect OCR affecting indexing
  - Browsing errors
  - Could not add materials and make fixes without taking the whole collection offline
- The decision was made to update to DLXS v.13 (the version we were using for our other text collections)

Cornell's Making of America

- Needed to convert 1,470 SGML files to XML. Formats were similar, but not quite!

- 963 images needed to be replaced in 66 volumes, 14 journals and 3 monographs

- 6 journals had severe indexing problems that needed to be fixed (for example: some issues of Atlantic Monthly were actually North American Review issues).

- Needed to decide how to set up MOA:
  - One collection with multiple XPAT dd files?
  - Multiple collections?
  - Multiple collections in a single group?

- Serials needed to have article information preserved

Cornell's Making of America

- MOA is our most used web service outside of the Physics arXiv with millions of hits/year.

- Many Civil War enthusiasts and Genealogists use the site and are very vocal.

- Original Middleware had features which we wanted to make sure were duplicated:
  ◦ Search by date over entire collection
  ◦ Search within a journal or monograph
  ◦ A Browse that listed all monographs AND all articles and authors (generated by software)

Cornell's Making of America

- In our conversion, we made the following choices:

  - Decided to make each journal a collection

  - Decide to make one collection of all of the monographs

  - Decided to take the 2 Official Civil War monographs (158 volumes) and make them part of one collection

- Added all of these to a group (Moagrp)

Cornell's Making of America

- Problems encountered:
  - Date searching across the entire Making of America didn't work.

    - Our Solution: Set up 2 new Journal Collections (Journals A-L and Journals M-Z). So, the user could do date range searches within these groupings of journals. So, even though we didn't have date-range searching across the entire MOA, a user could do date searching:
    - within journals
    - across the monographs
    - across the Civil War documents
    - across 2 groupings of journals

Cornell's Making of America

- Problems encountered:
  - PDF's didn't work.
    - Solution:
      - path to pdfopt in symlinks was wrong. That fixed one problem
      - Newer images needed to run through
        tiffcp -r -1 sourceimg destimg to remove multiple strips
  - Bookmarked links no longer worked (including Library of Congress' *American Memory Site*
    - Solution:
      - Created a series of Apache web redirects
      - Created scripts to redirect specific calls to articles and pages

Cornell's Making of America

- Problems
  - Some images were large and under tif2gif they were rendered large, but under tif2web they were smaller. Certain researchers complained
    - Solution: created static derivatives using tif2gif

  - Browse pages containing all monograph and article authors and titles could not be easily reproduced
    - Solution: using log analysis, we found that this specific feature in our original site was used by very few people, so we opted not to duplicate it

Browse a complete Bibliography of Books and Articles in MoA, organized by author and title

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | Non-alphabetic

Cornell's Making of America

- Problems
- The War of the Rebellion volumes are in the thousands of pages. Clicking on *"View the Entire Text"* results in an Internal Server Error.
- Solution:
  - No good solution
  - In some cases we create PDFs for users
  - Direct users to Internet Archive which has full-text

Cornell's Making of America

- New User Interface

Cornell's Making of America

- Some users *hated* the new site
- We received e-mails with comments like:

  ◦ What have you done?

  ◦ WHAT DID YOU DO.....to the SEARCH ENGINE??

  ◦ I want the old interface back!

  ◦ Where are the gathered records?????

  ◦ "…the new set up is bad, really, really, bad."
  ◦ A poor choise *(sp)* of formating *(sp),* not an improvement at all."

  ◦ "THIS NEW FORMAT MAKES THINGS VERY CONFUSING, TIME CONSUMING, and DISCOURAGES USE OF THE DATA BASE. Is this what you want??

  ◦ "WHO CARES WHAT THE SITE LOOKS LIKE??? Is this a beauty pagent or what?"

  ◦ "I am sorry I am being such a bitch about this, but it's all your fault!"

Cornell's Making of America

- In general, we have about 250,000 sessions per month.

- We received about 44 complaints concerning the new interface

- Interestingly enough once we answered and/or satisfied the users who complained, we received no further complaints until June of this year. It appears that some researchers only use the site once a year (Civil War Re-enactors?) and they want the site to remain the same.

- Lessons Learned:
  ◦ Log Analysis for underused features/functionality is a good place to start
  ◦ In future platform changes/redesigns we should do usability testing

- The upgrade has given us the flexibility to add to the collection and we have done so many times since then.

Local Modification (used across all collections)

- We have made 2 code modifications to our DLXS installation to handle static images. Both of these involve ItemView.pm

- General Static Images:
  - We place our static derivatives in a directory {'DLXSROOT'}/web/c/collid/pages

  - In ItemView.pm we look first in these directories which are explicitly defined in ItemView.pm. If the derivatives are found they are used, else we look in the cache and/or create the images on-the-fly

  - We have named these derivatives: image-s.gif,image-S.gif, image-l.gif and image.pdf.

Local Modification (for single static gifs)

- Our Witchcraft collections consists of very good derived text for each page but single gif images of one size.

- In response (and with the help of Phil Farber) we made a change to ItemView.pm in DeliverFileHandler to call 2 subroutines:
    - sub SetTargetImageFormat
    - sub GetTargetImageFormat

- Then in Collmgr we added the following for the Witchcraft Collection:
    - In *pvinputoutputmap*, we added "gif=gif"
    - In *pvstoredformatprioritymap*, we added "image=gif"