

Making of America - Phase I:

A Collaboration Between Cornell University & the University of Michigan

The Collection

- 1850 - 1877
- 1.5 million images when complete
- 5,000 images

Making of America (MOA) represents a major collaborative endeavor to preserve and make accessible through digital technology a significant body of primary sources documenting American social history from the antebellum period through reconstruction. When the initial phase of the project is completed, the MOA collection will include over 1.5 million images, representing approximately 5,000 volumes of primary source materials drawn from the depth of the Cornell and Michigan libraries. The Cornell selection process has focused on the major serials of the period, ranging from general interest publications to those with more targeted audiences. The process at Michigan has focused on monographs in the subject areas of education, psychology, American history, sociology, science and technology, and religion.

The Conversion Process

- 600 dpi
- TIFF images
- CCITT Group 4

The materials in the MOA collection are scanned from the original paper source. The images are captured at 600 dpi in TIFF image format and compressed using CCITT Group 4. Minimal document structuring occurs at the point of conversion, primarily linking image numbers to pagination and tagging self-referencing portions of the text. In the case of serials, low-level indexing will be added post-conversion by the partner institutions.

The Cornell Prototype

http://moa.cit.cornell.edu/MOA/moa-main_page.html

The interface offers basic search and retrieval of page images. In each case, the user decides on an item of interest by using an image of the journal index as it appears in the volume. The user can then move easily to the appropriate page image by means of a separate "turn to

- 400,000 images loaded to date

Classroom access prototype:

<http://moa.cit.cornell.edu/MOA/COURSES/classes.html>

The University of Michigan Prototype

<http://www.umdl.umich.edu/moa/>

- 183,000 images (682 volumes) loaded to date

page" button. Different image size options are provided, including thumbnails.

The separate classroom access prototype, which is presented in the frames version of Netscape, includes consolidated indexes for browsing each serial by author and by title; these indexes are hyperlinked to the articles. A search feature for the indexes is being developed.

Further conversion of the University of Michigan materials (approximately 650,000 pages within the total project) includes both optical character recognition of the page images, and SGML-encoding of the ensuing textual information. Assessment measures to evaluate the quality of the OCR are also being developed. If this process is successful, we will be able to prioritize MOA materials for a process where we might refine the OCR and SGML, allowing us to show the encoded text with links to images.

Developed by Digital Library Production Services (DLPS) staff, the MOA prototype interface utilizing low-level SGML mark-up offers: basic search and retrieval, with rendering of page images; a browse feature that lets users take a look at the variations in names, title words, and subject headings; a frames-based navigation feature for resizing pages to facilitate navigating through volumes online; and, a mechanism to link the local OPAC records to items in the MOA collection.

