



The Production Process

The conversion of the backfile archive of the scholarly journal literature available through JSTOR takes place at two sites:

- JSTOR Main Office, New York, New York
- JSTOR Production, Ann Arbor, Michigan

The process begins in New York, where negotiations between JSTOR's Director of Publisher Relations and journal publishers result in a non-exclusive license for JSTOR to undertake the conversion of a particular title from paper to digital format and then make it available to the scholarly community.

Physical copies of the title are first solicited directly from the publisher; however, in the case of missing or damaged issues, JSTOR purchases copies from a variety of periodical replacement services. JSTOR also accepts loans or donations of [back issues needed](#) from a variety of sources, including JSTOR participating institutions.

These copies arrive in Ann Arbor, where JSTOR production staff inventory them to confirm that a complete run of the title exists. Through a page-by-page examination of each issue, a publication record is created for the title. Preservation concerns are addressed during this part of the process and scanning guidelines are created. A production librarian, a serials specialist on the JSTOR staff, examines the structure of each title and creates appropriate indexing guidelines, matching the indexing specifications used by JSTOR to each individual title and article. Then, each journal title is shipped to a contractor for scanning and data entry.

At the contractor's facility, the physical journal volumes are disbound and separated into discrete issue units. Each page is scanned at 600 dpi resolution, with meticulous attention to quality. Page images are checked for marks, folds, or skewing, and are rejected if deemed unacceptable.

Subsequently, a table of contents file, which includes bibliographic citation information and an item type identifier (full length article, book review, advertisement, etc.) — as well as key words and abstracts if these exist in the original publication — is keyed for each article in the journal run. The indexing guidelines, created for each title by JSTOR production librarians, ensure that the components of this table of contents file will facilitate more effective searching and browsing in the digital environment. All digital files created by the contractor — the page images and the table of contents files — are downloaded to CD-ROM for shipment back to the JSTOR production facility.

The final part of the JSTOR production process again happens at the JSTOR production facilities, where files are uploaded from CD-ROM to the JSTOR file servers. An intensive quality control process verifies both the quality of the page images and the accuracy of the information contained in the keyed table of contents files. Once an initial quality control check has been performed and the data judged acceptable, each page image is then processed by optical character recognition (OCR) software in order to create full-text for searching. Generally, the OCR program yields a 97% accuracy rate, with some journals having OCR files that are 99.95% accurate, or about one error in every 2,000 characters. Although this level of accuracy is not high enough for presentation to users, it has proven to be quite satisfactory for searching.

When further quality control reviews have been completed and the files have been compressed using [Cartesian Perceptual Compression](#), the availability of the title is announced to JSTOR participants.

Last updated February 9, 2004.

