Why Images?

As JSTOR's Mission & Goals suggest, we strive to meet many objectives and to satisfy the needs of all participants. The original concept for JSTOR was to convert the back issues of paper journals into electronic formats that would allow for savings in space (and in capital costs associated with that space) while simultaneously improving access to the journal content. Thus, it is equally as important for JSTOR to be providing faithful replications of the original print journals as it is for it to be providing access to the archive, since the electronic version is to be used as a substitute for the print version.

One important technological decision that JSTOR made was to deliver the content of the archive as images. We decided to combine the advantages of page images with a searchable text index, and JSTOR stores the data in both forms. JSTOR delivers scanned page images to its users, while using the raw text files (created using Optical Character Recognition (OCR) software) behind the images for search purposes.

Benefits of Images:

- **Faithful Replication:** If, in keeping with our mission to function as a trusted archive, JSTOR is to serve as a substitute for the journal volumes on the shelves, it must offer an electronic version that is a faithful replication of the original. An image-based approach ensures the integrity of the materials in the archive, while also retaining the appearance and "look and feel" of the journal in its original presentation. This is central to our mission and a key basis upon which JSTOR was founded.

- **Representation of Non-Text Content:** Whether they appear as photographs, charts, tables, or special characters and formulae, certain components of articles generally cannot be displayed with 100% accuracy using text-based methods available to standard web browsers.

- **Accuracy of Images:** Page images are 100% accurate. JSTOR creates a text index for search purposes as part of its production process through the use of OCR software. Our scanning vendor conducted a series of reviews of OCR samples on a variety of materials and found a 97% average accuracy rate (on uncorrected text). In JSTOR, some journals will have OCR accuracy rates as high as 99.95%. But, although our OCR is accurate for search purposes, it is unacceptable for display, owing to typographical, word order, formatting, and other elements that are not accurately represented. The appearance of typographical and other errors could undermine the perception of quality that publishers have worked long and hard to establish and that users of all kinds expect. Indeed, while users in the visually impaired and learning disabled communities might prefer text, displaying our OCR would not offer a product and experience that is equivalent to what users in the non-visually impaired and non-learning disabled communities encounter. Appropriate assistive technology designed specifically for the visually impaired and learning disabled communities can offer far better accuracy than JSTOR can were we to display the OCR'd text we have created for search purposes.

The importance to libraries and publishers, as well as to the fulfillment of our not-for-profit mission, of faithful replications of journals, the ability to display non-textual material accurately, and the problems associated with displaying the OCR text we have created for search purposes are the primary motivations for JSTOR's use of images as the mechanism for delivery of journal articles.

We are aware that our image based approach causes certain difficulties for users who are visually impaired or learning disabled and use assistive technologies to access material on the Internet. JSTOR now offers options to help alleviate some of these difficulties. For more information, please see JSTOR and Accessibility.

Last updated July 15, 2003.