

**KILPATRICK TOWNSEND & STOCKTON LLP**

Joseph Petersen (JP 9071)  
Robert Potter (RP 5757)  
1114 Avenue of the Americas  
New York, NY 10036  
Telephone: (212) 775-8700  
Facsimile: (212) 775-8800  
Email: jpetersen@kilpatricktownsend.com

Joseph M. Beck (admitted *pro hac vice*)  
W. Andrew Pequignot (admitted *pro hac vice*)  
Allison Scott Roach (admitted *pro hac vice*)  
1100 Peachtree Street, Suite 2800  
Atlanta, Georgia 30309-4530  
Telephone: (404) 815-6500  
Facsimile: (404) 815-6555  
Email: jbeck@kilpatricktownsend.com

*Attorneys for Defendants*

**UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK**

THE AUTHORS GUILD, INC., ET AL.,

Plaintiffs,

v.

HATHITRUST, ET AL.,

Defendants.

Case No. 11 Civ. 6351 (HB)

**SUPPLEMENTAL DECLARATION OF JOHN WILKIN IN SUPPORT OF  
THE LIBRARIES' MOTION FOR SUMMARY JUDGMENT**

I, John Wilkin, pursuant to 28 U.S.C. § 1746, hereby declare as follows:

1. I understand that the Plaintiffs, in their Opposition to the Libraries' motion for summary judgment, have questioned the Libraries' use and retention of image and text files in the HathiTrust Digital Library ("HDL"). As discussed below, image and text files of each work are necessary for the search, preservation, and accessibility services HathiTrust provides.

2. The digital copy of each work in the HDL includes (a) an image component representing photographic reproductions of pages of the work (the “Image File”) and (b) a Unicode text component representing text in machine-readable format (the “Text File”). The Text File is created from the Image File using Optical Character Recognition (OCR) software that converts the page images into searchable text.

3. Maintaining only the Image File, or only the Text File, would not permit HathiTrust to provide its search, preservation, and accessibility services. For example, the Image File preserves for replacement purposes the text, formatting, images, and other features on the page as they appear in the book, but it cannot provide full-text searching. Conversely, the Text File, which allows full-text searching, cannot serve as an archival preservation format.

4. First, the Text File does not include completely accurate text. Even the best OCR technology available does not reliably recognize all characters correctly, particularly in the case of older or inconsistent fonts or creative typography. For example, “L’s” often become “1’s” and “s’s” in older fonts are often incorrectly identified by OCR software as “f’s.”

5. Second, existing OCR software is not capable of producing a Text File that includes all of the textual, formatting, and graphical features of a book. Through manual XML coding, we are able to identify and describe certain textual features, but running heads (a short title that appears at the top of each page), paragraphs, stanzas, and line breaks are either not coded or are not reliably included in coding. Moreover, illustrations, tables, graphs, and other images are not included in the Text File. These textual, formatting, and graphical features missing from the Text File may represent information necessary to communicate the information in the work. For example, in poetry and other creative writing forms, paragraph or stanza format, layout, and line breaks may be relevant to the works’ meaning. In addition, works that include

mathematical or scientific formulas often rely on superscript and subscript notations and other positional relationships between characters and symbols that are not reliably identified by OCR software, and maintaining a Text File alone for these works would be insufficient.

6. Because the Text File does not include all of the necessary information, as described above, the Image File remains the authoritative digital representation of the printed book. The Image File may also be used to improve the accuracy of the Text File as OCR technology enhancements becomes available.

7. Moreover, both the Image File and the Text File are critical to HathiTrust's fulfillment of its mission to provide equal access to users with print disabilities. Some blind users may be able to utilize a text-only digital format by using screen-readers and text-to-voice software that convert the text into an accessible format. Other print-disabled users—such as low-vision readers or sighted individuals with other print disabilities—may be able to read a digital image file that has been enlarged or otherwise optimized for their use. Providing these users with a text format only would deny them the ability to access information communicated in a book's text formatting and layout, special symbols or characters, or graphical features such as photographs, illustrations, graphs, or tables. Only by making the Image File available to these users can HathiTrust provide access more equivalent to that of their peers without print disabilities.

8. Recognizing that print disabilities take a variety of forms and that individuals with different print disabilities may require different formats, HathiTrust offers students and faculty with certified print disabilities both the Image File and a concatenated presentation of the Text File that is optimized for use with screen-readers and text-to-speech software.

I declare under penalty of perjury under the laws of the United States that the foregoing is true and correct.

Date: July 26, 2012



John P. Wilkin