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

I. Qualifications and Statement of Assignment

1. I am the Frederick R. Kappel Chair in Applied Economics at the Carlson School of the University of Minnesota and a Research Associate at the National Bureau of Economic Research. I received a Ph.D. in Economics from Stanford University in 1990. Prior to joining the faculty of the Carlson School, I was a professor at the Wharton School of the University of Pennsylvania, where I was the Joel S.
Ehrenkranz Family Professor of Business and Public Policy. From 1990 to 1997 I was an assistant professor and later an associate professor of economics at Yale University. I have authored and/or edited four books and published over 40 articles in economics journals. My main areas of research are industrial economics and law and economics. Within industrial economics I have published articles on many aspects of media industries, including music, movies, radio, television, newspapers, and the Internet. Much of my research involves studying the development of new types of online services that make use of existing products or services, and assessing their impact on the market for those existing products.

2. I have twice served as a consultant to the Federal Communications Commission on issues related to media ownership, from 2001 through 2003 and again from 2010 to 2011. I have also written on economic issues for popular audiences, as the Dismal Scientist columnist for Slate Magazine from 2006 until 2009 and as the author of Scroogenomics (Princeton University Press, 2009).

3. My curriculum vitae is attached as Appendix A. A list of cases in which I have testified in the past four years is listed in Appendix B.

4. I have been asked by Counsel for the HathiTrust and the defendant libraries to determine whether certain services provided by the HathiTrust—specifically the ability to search digitized books—could be undertaken as a viable and licensed commercial enterprise. I have not analyzed whether a different type of offering, providing different or additional services, would be commercially viable, and have limited my inquiry only to the viability of a service like the HathiTrust. To answer this question I first compare the costs that a commercial entity providing similar services would incur against the potential revenues that such an entity could expect to
earn. Second, I compare the economic role of such a hypothetical service to existing, analogous services, in light of economic research on the relationship between the availability of information through search services and the demand for the underlying indexed works.

5. This report and the opinions expressed herein are based on my analysis of the materials reviewed to date, together with my training, education, and experience. The documents I have reviewed and data I have analyzed in preparing this report are listed in Appendix C. I reserve the right to supplement my conclusions should additional documents, testimony, or other materials become available to me.

6. I am being compensated at my hourly billing rate of $550 per hour for my work in this matter. Payment is not contingent on my opinion expressed, nor on the outcome of this case. Part of the work on this matter has been performed under my direction by economists at The Brattle Group, Inc.

II SUMMARY OF CONCLUSIONS

7. Based on my analysis, I have reached the following conclusions:

- The costs associated with creating and maintaining a service like the HathiTrust would exceed any potential revenue that such a venture could earn. Thus, I conclude that the creation and offering of a service with the functionality of the HathiTrust, but with licensed content, is not a commercially viable endeavor (i.e., an endeavor where the revenues could cover the costs).

- In addition to incurring the direct costs associated with creating and maintaining a database like the one available through the HathiTrust, an entity creating such a service, were it to license the use of every copyrighted work, would have to
expend tremendously significant resources to secure the rights to works contained in the HathiTrust.

- The potential revenues associated with such a venture would not be sufficient to make a licensed searchable database commercially viable.
- Research shows that services offering close substitutes for existing works (such as pirate music or movie sites) cannibalize demand for the underlying works. By contrast, research shows that services offering incomplete elements of existing works can stimulate demand for those works. Comparing the HathiTrust’s service with recent research on the demand impact of piracy and related phenomena, I conclude that the HathiTrust service is likely to stimulate demand for the works searchable in it.

III  DATA

8. To conduct my analysis, I estimate the cost of creating a searchable database similar to the HathiTrust’s, but with licensed content, and I compare this against an estimate of the revenue that such a database could generate.

9. For my analyses I obtained data on advertising revenue at Facebook and Twitter, two of the most visited websites in the world. I also collected data on web traffic for those sites as well as for the HathiTrust web site and other book-related websites, from the website Alexa (www.alexa.com).
IV  ANALYSIS

A.  Costs Associated with Creating a Licensed Database Like the HathiTrust’s

1.  Direct Costs to Scan Works

10.  There are two broad types of costs that a commercial entity would need to incur in order to establish a searchable database similar to that offered by the HathiTrust. First, the entity would face the costs of obtaining digital copies of each work in the database. Creating a database of the size and scope offered by the HathiTrust would require, in almost every case, both acquiring a print copy of the work and then literally having it scanned. Second, I understand that Plaintiffs believe that the entity would also need to incur the costs of obtaining permissions to include certain of the works in the collection.

11.  For some works, notably those for which a digital file already exists, the cost of scanning is likely to be low. Most works created in the recent past, or those that will be created in the future, will already have a digital file that could be imported into the entity’s database, assuming the entity could gain access to the digital file. However, for works for which the entity would not be able to access an existing digital work, the work would have to be scanned. To date, the HathiTrust’s digital collection contains 10.4 million scanned volumes.\(^1\) I have been advised that, because the overwhelming majority of these volumes are older works that were not “born digital,” very few of the relevant volumes currently in the HathiTrust’s database would be

\(^1\) HathiTrust, http://www.hathitrust.org/.
available to the entity in electronic form, so my basic calculation assumes that all works would require scanning.²

12. I conducted research on the cost of large-scale scanning. The largest-scale scanning project that I am aware of is the Google Books project, which has digitized approximately 20 million books as of March 2012.³ Google estimates its cost simply to scan a given book, not including other costs associated with the project, to be approximately $4.⁴ Google further estimates that it has spent approximately $3 million to date for the operational cost to scan a book (including labor and equipment), maintaining books after being scanned, and technology development.⁵ Thus, Google estimates its overall average cost has been approximately $3 per book. Because the Google Books project is the largest scanning project that I am aware of, I will use its estimate of scanning costs in my estimate of scanning costs. As shown below, it is also the lowest scanning cost of any estimates that I could find. Using the Google estimate therefore provides a conservative estimate of the scanning costs.

13. Another large-scale book scanning project is the Million Book Project, a large-scale project that, as of December 2007 had scanned more than 1.5 million volumes.⁶ The

---

² As I discuss below, see ¶¶40-44, my conclusions are unaffected by the possibility that some works—or even all of the works—would not need to be scanned. Even if scanning costs were zero, I find that the other costs associated with creating a licensed database of this scope exceed any revenues that such a service could be expected to generate.


⁵ Deposition of Daniel Clancy, June 1, 2012, p. 52:10-16.

project was supported by the National Science Foundation, which has provided $3.6 million\(^7\) in “funding for equipment.”\(^8\) In addition, the Project reports that manpower provided by India and China “represent[] a twenty-to-one relative contribution … to this project.”\(^9\) This suggests that the Million Book Project has expended resources of $75.6 million, or approximately $50 per book scanned.

14. A 2006 *New York Times* article on the Million Book Project reported on the efforts to scan 30,000 volumes from the Carnegie Mellon University and the Carnegie Libraries. The initial scanning cost was reported to be $10 per volume.\(^10\) This appears to understate their actual costs, as the project seeks donations and currently reports an average scanning cost of approximately $20 per book.\(^11\)

15. I also conducted a search of commercial services that offer to scan books. Such services quote prices depending on certain factors, including whether the scan is destructive or non-destructive (i.e., whether the book is to remain intact after scanning). Prices advertised on the Internet by services offering scanning run as low as $6.95, plus 5 cents per page, for a destructive scan.\(^12\) The average volume in the HathiTrust is 350 pages, so to scan it destructively through this service would cost


\(^11\) The Million Book Project’s website reports that a $1,000 contribution “permits the digitization of about 50 books.” Larger contributions are also reported to permit digitization at approximately $20 per book. For example, the website notes that a $100,000 donation permits 5,000 books to be digitized. See http://www.ulib.org/ULIBDonations.htm (accessed June 15, 2012).

\(^12\) http://boundbookscanning.com/ (accessed June 18, 2012). Bound Book Scanning also offers non-destructive scanning for $8.95, plus 8 cents per page, with restrictions on the size and thickness of the work. The cost of scanning a 350-page book (the average length of books in the HathiTrust) using the non-destructive method would cost $36.95. To evaluate the relative benefit of destructive versus non-destructive scanning, one would have to include the cost of the book destroyed through destructive scanning, assuming that a replacement copy would even be available.
$24.45, (plus the cost of obtaining a replacement copy of the work, including shipping, as well as other administrative costs).\(^{13}\)

16. For the remainder of the report, I will use the Google Books projects cost of $\_\_\_ per work as the estimated cost a commercial entity would incur when scanning millions of works into a database. The Google Books estimate is the lowest scanning cost that I have found, and I understand they can be much higher. Using this estimate, the resulting scanning-only cost of producing a commercial HathiTrust-like service is 10.4 million x $\_\_\_, or $\_\_\_ million.

2. *Indirect Costs of Securing Copyrights*

17. It is my understanding that Plaintiffs in this case believe that an entity developing a service similar to that the HathiTrust would have to secure the rights to certain of the works to make them available for search. Therefore, I have investigated the costs associated with securing those rights. I am unaware of systematic quantitative evidence on the cost of securing such rights, although numerous narrative accounts exist (see below). These accounts make clear that securing the rights would be a manual, labor-intensive process that would, in most cases, require contacting each rights holder individually for permission. I have been advised that the Copyright Clearance Center, a collective rights organization that collects royalty payments for written works and distributes them to rights holders, does not offer licenses for searchable databases of textual works, nor does it currently have any plans to offer such licenses.\(^{14}\)

\(^{13}\) The HathiTrust has scanned 10.4 million volumes, for a total of 3.641 billion pages, which averages to 350 pages per volume. See http://www.hathitrust.org/ (accessed June 26, 2012).

\(^{14}\) See Deposition of Frederic L. Haber, June 4, 2012, pp. 19-32.
18. From 2006 through 2008, the United States Congress studied the issue of orphan works. During the investigation, the U.S. Copyright Office requested comments from interested parties. Several university libraries submitted comments detailing their attempts to secure copyright permissions to digitize orphan works. For example, Carnegie Mellon University submitted a letter to the Copyright Office detailing attempts to determine the feasibility of identifying copyright holders for works to be scanned. The letter explained the difficulty in determining the copyright holder of numerous works. The university could not locate publishers for 22 percent of the books in its sample. In addition, 36 percent of those publishers that were identified did not respond to multiple requests by the university to secure permissions. Thus, Carnegie Mellon did not receive any information from the publishers for half of the works for which it attempted to secure permissions. Overall, Carnegie Mellon estimated the direct administrative cost of receiving permission was $78 per title for which permission was granted. Some copyright holders then requested fees ranging from $50 to $300 to grant such permissions. The Carnegie Mellon letter notes that these costs are conservative because they do not include costs associated with consulting the University’s legal counsel, creating a database to track progress of their efforts to secure permissions, or paying the staff needed to manage the project.

19. Cornell University also submitted a letter detailing a project to digitize 343 monographs that were out of print but under copyright. Cornell could not identify the

15 Carnegie Mellon identified publishers for 78 percent (100-22) of works in its sample. It received responses from 64 percent (100-36) publishers. These figures indicate that the university received responses for 50 percent (0.78*0.64=0.499) of the works for which it sought permission.

copyright holders for 198 of these works (58 percent). The university reported that it spent over $50,000 in staff time and was able to receive permission for only 98 works. Thus, the average cost of securing permission was $510 per work whose permission was successfully secured.  

20. Similarly, Brigham Young Law School submitted a letter detailing unsuccessful efforts to secure rights to use certain works in courses. In one example, in 2000, the Law School spent “over a dozen hours” working to acquire copyright permission for a single work. In another instance, in 2001, the Law School spent “upwards of twenty hours” trying to determine the copyright holder of a work whose author was deceased and whose publisher was no longer in business.

21. In sum, the evidence from the comments submitted to the Copyright Office indicates that securing rights for individual works is a manual, time-consuming process, and that the cost of securing those rights appears to be significantly higher than the cost of scanning the works.

22. Two things appear clear from the available evidence concerning the challenges associated with obtaining permissions. First, for works where permissions can even be obtained, the cost of obtaining such permissions is significant. The estimates above place this cost between $78 to $510 per work, and this does not include any licensing fee or royalty that may be demanded by the identified rights holder. Second, finding the rights holder appears to be entirely infeasible for a large share of


works, with more than half of all such holders remaining unidentified or unconfirmed (half or more in the experience of both Carnegie Mellon and Cornell.).

23. An entity seeking to duplicate the HathiTrust’s database would not need to secure permissions for all 10.4 million works because some are already in the public domain. The HathiTrust reports that 3.1 million works in its collection are in the public domain.\(^{19}\) Therefore, the entity would have to obtain rights for 7.3 million works.\(^{20}\)

24. While it is difficult to translate these findings into simple estimates of cost, a few calculations are economically relevant. Utilizing the lowest identified estimate as to the costs of obtaining permissions ($78 per work), and applying this to the estimated 7.3 million copyrighted works in the database, leads to a total “permissions cost” of $569 million. (This assumes that the entity could actually identify and obtain permissions from every rights holder; as noted above, Carnegie Mellon and Cornell were not even able to identify the majority of relevant rights holders.) Thus, the overall cost of creating a licensed database like the one offered by the HathiTrust—including obtaining permissions for all the works for which Plaintiffs claim permissions are required—comes to a total of $\_\_ million (this includes both the estimated $\_\_ million in scanning costs discussed in paragraphs 10 through 16, as well as the $569 million needed to obtain permissions from rights holders, but does not include the payment of any royalties or licensing fees to rights holders).

\(^{19}\) See http://www.hathitrust.org/ (accessed June 18, 2012).

\(^{20}\) I have been advised that some of these 7.3 million works, published between 1923 and 1963, are likely already in the public domain, but that making that determination would itself be an intensive and expensive process. Accordingly, I understand that the HathiTrust treats all of these works as if they are protected by copyright.
25. Going forward, I will base my calculations on this estimated $\_\_\_\_ million total cost. As I will show below, no entity creating and offering such a database could generate enough revenue to cover these costs. As I explain further below, my conclusions do not change even if I make a wide range of alternative assumptions about such costs.

B. Potential Revenue Associated with a Searchable Database

26. As discussed above, an entity creating a service like the HathiTrust would incur $\_\_\_\_ million in up-front scanning and permissions costs to make the database available for search. To be commercially viable, the service would have to earn sufficient revenues to cover those costs. Because the revenues would be earned over time, it is necessary to account for this by calculating a present discounted value of future revenues.\(^{21}\) I calculate the present value of revenues using a discount rate of 10 percent per year. This discount rate is approximately equal to the cost of equity for a firm in a related line of business, Reed Elsevier, an academic publisher that publishes over 2,500 academic journals that are searchable using its ScienceDirect service.\(^{22}\) Given the $\_\_\_\_ million cost to scan 10.4 million volumes and obtain the necessary permissions for those works not in the public domain, any service offering a database like the HathiTrust’s would need to generate approximately $\_\_\_\_ million in annual


\(^{22}\) Reed Elsevier is a Dutch firm with American Depositary Receipts that are publicly traded on the New York Stock Exchange. Bloomberg reports Reed Elsevier’s cost of equity to be between 10.3 and 10.5 percent (Accessed June 27, 2012). As a comparison to an internet search firm, Bloomberg reports Google’s cost of equity to be 10.8 percent.
revenue to cover this cost. As I discuss below, a licensed service similar to the HathiTrust would not be able to generate this level of revenue.

27. A commercial entity offering a searchable database like the HathiTrust’s could potentially generate revenue in two ways: through an ad-supported website and by making the database available to paid subscribers. My conclusion is that neither of these methods, whether alone or in combination, would generate sufficient revenue to make such a database commercially viable.

1. Ad-Supported Site

28. A HathiTrust-like commercial service might support itself with ad revenue. Given information on the relationship between site traffic and ad revenue generally, along with information suggesting the site traffic that a commercial HathiTrust-like site would realistically attract, it is possible to calculate the ad revenue that a HathiTrust-like site could feasibly generate.

29. Advertising revenue data for websites is not generally available, but it is possible to get revenue data for certain specific sites. In 2011, for example, Facebook generated $3.1 billion in ad revenue, and Twitter generated $140 million. Alexa.com collects information on website traffic, including information on “reach” (a measure of the

---

23 The present value of earning one dollar every year in the future at a discount rate of \( r \) percent is equal to \( 1/r \). Thus, if \( r=10 \) percent, the present value of earning one dollar every year is \( $10 (1/0.1=10) \). (See Richard A. Brealey, Stewart C. Myers, and Franklin Allen, *Principles of Corporate Finance* (9th Ed.) Boston: McGraw-Hill Irwin, 2008, pp. 40-41 for a more detailed discussion.)

24 My conclusions would be unaffected with the use of lower discount rates. For example using a risk-free discount rate of 4.58 percent (as reported by Ibbotson), the service would not cover its costs. However, it is be incorrect to use a risk-free discount rate because the cash flows associated with the service would be subject to risk. (See Morningstar, Ibbotson Cost of Capital 2010 Yearbook, Chicago: Morningstar, Inc. (2010), p. 23.)

number of users)\textsuperscript{26} and time on site (estimated daily time on site). Hence, Alexa’s 3-month reach measure indicates that 44.7 percent of global Internet users visit Facebook each day, on average, in the last three months. Similarly, Twitter’s 3-month reach measure of 9.2 percent indicates that 9.2 percent of users visit Twitter each day, on average. Based on the time-on-site measure, Facebook users spend an average of 1,421 seconds on that website each day, averaged over the past 3 months, while Twitter users spend an average of 429 seconds daily on that site. The product of reach and time at site provides a measure of overall time spent at the site, added across Internet users, which I term “reach seconds.” By this measure, Facebook has 636 reach seconds per day (meaning that the average amount of time at the site is 636 seconds per day, averaged across all Internet users, including both those visiting Facebook and those not visiting Facebook). The analogous measure for Twitter is 40 seconds per day. Dividing annual revenue for Facebook by its reach seconds yields an estimate of $5 million per year per reach second, while the analogous figure for Twitter is $4 million.

30. These estimates of revenue per audience measure provide a way to measure the level of advertising revenue that a HathiTrust-like website could generate. HathiTrust.org’s reach measure is 0.00155 percent, which is 1/28,000 as large as Facebook’s reach of 44.7 percent cited above. HathiTrust.org’s time on site averages 292 seconds, so its reach-second measure is 0.00453, which is less than 1/140,000 as

\textsuperscript{26} According to Alexa.com, “Reach measures the number of users. Reach is typically expressed as the percentage of all Internet users who visit a given site. So, for example, if a site like yahoo.com has a reach of 28%, this means that of all global Internet users measured by Alexa, 28% of them visit yahoo.com. Alexa’s one-week and three-month average reach are measures of daily reach, averaged over the specified time period. The three-month change is determined by comparing a site’s current reach with its values from three months ago.” See http://www.alexa.com/help/traffic-learn-more (accessed June 26, 2012).
large as Facebook’s reach second measure of 636 seconds cited above. Hence, based on Facebook’s ad revenue per reach second, I would expect a HathiTrust-like website to be able to generate 1/140,000 of the $3.1 billion in annual ad revenue that Facebook generates, so such a site would generate $23,000 in annual revenue.\(^{27}\) HathiTrust.org generates a reach second measure 1/8,700 as high as Twitter’s 40 seconds cited above. Hence, based on Twitter’s revenue per reach second, I would expect a HathiTrust-like commercial site to be able to generate 1/8,700 as much revenue as Twitter, or $18,000 in annual ad revenue.\(^{28}\)

31. The existing HathiTrust website is not the only book-oriented site on the Internet, and so it is instructive to use the above methodology to estimate the ad revenue that other book-related sites could generate. For example, the Project Gutenberg site (gutenberg.org), which distributes public domain books free of charge and has far more traffic than the HathiTrust site, has a 3-month reach of 0.0228 percent and an average time on site of 268 seconds. Using the method employed above, I estimate that a site with this level of traffic could produce annual ad revenue between $244,000 and $306,000, based on Facebook and Twitter’s revenue per reach second, respectively. Using the same approach for other book-related sites—AbeBooks, JSTOR, and WorldCat—gives respective ranges of $360,000 to $452,000 for AbeBooks; $267,000 to $335,000 for JSTOR; and $153,000 to $193,000 for WorldCat. In the interest of making my calculations as conservative as possible, I adopt the largest of these estimates and therefore conclude that a HathiTrust-like site

---

\(^{27}\) HathiTrust.org reach (0.00155 percent) multiplied by its time on site (292 seconds) and Facebook’s estimated $5 million per reach second yields an estimated ad revenue of $22,630.

\(^{28}\) HathiTrust.org reach (0.00155 percent) multiplied by its time on site (292 seconds) and Twitter’s estimated $4 million per reach second yields an estimated ad revenue of $18,104.
could generate no more than $452,000 per year in ad revenue (i.e. the amount earned by AbeBooks, which has far higher website traffic that the HathiTrust).

32. The ad revenue that a HathiTrust-like commercial site could generate (no more than $452,000 per year) falls far short of the annual cost of operating such a site ($ million). I conclude that such a site would not be able to cover its costs with ad revenue.

2. **Subscription Service**

33. A commercial entity could also potentially charge libraries and other users subscription fees for the ability to search. How much could a private entity earn selling libraries subscriptions to a searchable database like the one offered by the HathiTrust? Analysts have identified four major products for digital discovery. While none is directly comparable to a hypothetical commercial service resembling the HathiTrust, one of them is sufficiently similar to support an instructive comparison. The WorldCat Local product includes a total of 969 million items, including the very works within the HathiTrust:

674+ million articles with one-click access to full text, 29+ million digital items from trusted sources like Google Books, OAIster and HathiTrust, 13+ million eBooks from leading aggregators and publishers, 44+ million pieces of evaluative content (Tables of Contents, cover art, summaries, etc.) included at no additional charge, 221+ million books in libraries worldwide.

Of these items, at least 703 million (674 million + 29 million) include some full-text functionality. For example, 674 million include “one-click access to full text,” while


others (like copyrighted works from the HathiTrust) include full text searchability, although with no access to full text.

34. HathiTrust works thus make up 1.48 percent (10.4 million/703 million) of the works within WorldCat Local that have some degree of full-text functionality. For the sake of discussion, if we ignore other offerings in WorldCat Local and, hypothetically, attribute all of its revenue-generating capability only to the works with full-text functionality, then we could infer that the HathiTrust works on their own could generate, at best, 1.48 percent of the total revenue that WorldCat Local generates. This is an overstatement in that it ignores the value of WorldCat Local’s other offerings, such as its information on “221+ million books.” It is potentially an overstatement in a second sense: the calculation presumes that the value of a service like WorldCat Local is directly proportional to the breadth of its collected works or, in other words, that two services with half the breadth of WorldCat Local’s collection would each be half as valuable as WorldCat Local. It is likely that this is not the case, and that the broader appeal of larger collections would generate disproportionately more revenue than two smaller ones.

35. How much does WorldCat Local generate in revenue? According to a published report, “Prices generally range from $9,000 to $25,000 per year for the subscription fee, depending on the size of the institution.” Calculating the revenue to WorldCat Local also requires an estimate of its sales penetration. According to an account in

---

Applying the pricing data cited above (a range of $9,000 to $25,000 per subscription) across 1,419 subscribers indicates a revenue range of $12,771,000 to $35,475,000 for a product that includes, as roughly one percent of its total content, all of the works in the HathiTrust.

36. We can narrow this revenue estimate somewhat further, however. The Library Journal account notes that WorldCat Local’s producer OCLC has “over 1200 employees and total revenue of $228 million in FY09/10. Yet only a relatively small portion of the whole relates to the types of products covered in this report; OCLC’s most recent annual report indicates only 7.6 percent of revenue, or around $17.3 million …” is attributable to services like WorldCat Local.

37. Accordingly, if OCLC generates $17.3 million in annual subscription revenue for searchable database products related to WorldCat Local, a commercial entity offering a database like the HathiTrust’s could generate annually only 1.48 percent of that revenue, or $256,000. This translates into approximately $180 in revenue from each subscriber to the HathiTrust-like database, assuming it had the same number of subscribers as WorldCat Local.

38. The current number of WorldCat Local subscribers (1,419) may actually understate the revenue available from future potential subscribers. According to OCLC, the

---

33 “OCLC is a nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world’s information and reducing information costs. More than 72,000 libraries in 170 countries and territories around the world have used OCLC services to locate, acquire, catalog, lend and preserve library materials.” See http://www.oclc.org/worldcatlocal/default.htm (accessed June 25, 2012).
entity producing WorldCat Local, their “core market is 4,000 academic and academic research libraries. These libraries also represent the core market for our publisher partners.” If an entity were to sell a HathiTrust-like services to all 4,000 of these libraries, at a subscription rate of $180 per year, this would generate $720,000 in total subscriptions.

There are 3,827 academic library systems in the United States. There are an additional 4,617 administrative units for academic libraries in the enlarged European Union. Thus, there are a total of 7,995 possible adopters in a set of countries that account for 48 percent of the world’s gross domestic product (GDP). Scaling this to the entire world suggests there are at most 16,656 libraries that might purchase a subscription to a searchable database like the HathiTrust’s. Even if every single one of these libraries paid $180/year to subscribe to a searchable database like the HathiTrust’s, this would produce only $2,998,080 million in total annual revenue.

C. Conclusion about Costs and Revenues

Above I estimate that the creation of a HathiTrust-like commercial service would cost $ million ($ million in scanning costs and $569 million in costs for

37 See David Fuegi Martin Jennings, “International library statistics: trends and commentary based on the Libecon data. Library services statistics in Europe and beyond,” LIBECON. 30 June, 2004. The enlarged EU includes: “Austria, Belgium, Denmark, Finland France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands Portugal, Spain, Sweden, UK and 10 countries were scheduled to join on 1st May 2004, namely Cyprus, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, the Slovak Republic, and Slovenia.” See Fuegi and Jennings (2004), page 8.
permissions). Given a 10 percent discount rate, such a service would need to generate more than $\text{[Redacted]}$ million per year in revenue to be commercially viable. I estimate that, even if deriving revenue from both advertising and subscription services, such a service could realistically hope to generate no more than roughly $3.5$ million in annual revenue, leaving a $\text{[Redacted]}$ million shortfall, and so I conclude that such a service would not even be close to viable.

41. As discussed above, my conclusion remains unchanged even when applied to a large range of possible alternative assumptions. For example, as indicated above, it is possible that an entity undertaking the creation of a licensed HathiTrust-like commercial service would not need to scan all $10.4$ million works. My conclusion is unaffected by this possibility: even without any scanning costs, the project would still cost $569$ million just for obtaining permissions (not including any actual license fees or royalties). Based on those costs alone, the service would need to generate roughly $57$ million in annual revenue for viability, which is still significantly far above my estimates of any realistic revenue expectations.

42. My conclusion also remains the same even if I were to apply a discounted value of future revenues lower than the 10 percent rate I used. With a lower discount rate, a searchable database service would need to generate less annual revenue to meet its costs. For example, discounting at the inappropriately low risk-free rate of 4.58 percent\textsuperscript{39} implies that, with an upfront cost of $\text{[Redacted]}$ million, the service could be viable with $\text{[Redacted]}$ million in annual revenue. This, again, far exceeds my estimate of the revenue the service could earn. Indeed, even if one assumes zero scanning costs

and an inappropriately low discount rate, the service would still require $26.1 million in annual revenue for viability. This too far exceeds my estimate of the revenue the service could generate.

43. I can examine the robustness of my conclusion in another way. Using the baseline estimates, the service would require $\[\text{millions}\] million in annual revenue for viability, while it would earn only $3.4 million. How much higher would revenue need to be for economic viability? Revenue would cover costs only if it exceeded my estimate by a factor of $\left(\frac{\text{millions}}{3.5}\right)$. Even if I calculate costs based on zero scanning costs and an inappropriately low risk-free discount rate, revenue would still need to be 7.5 times higher than my estimate ($26.1/3.5$) for the service to be viable.

44. For all of the reasons above, and even under a variety of alternative assumptions, I conclude that an entity seeking to create and market a licensed database like the HathiTrust’s would not be able to generate revenue sufficient to cover its up-front costs, and is thus not commercially viable.

D. Substitutes, Complements, and the Economics of Search and Discovery Tools

45. Since the development of digital technology for copying, many new products and services related to digitizable content (music, text, video) have emerged. When these new technologies have first appeared, it has not always been clear whether the new offerings function as substitutes or complements for existing products. That is, it is not clear whether they cannibalize or stimulate the demand for existing products. To take one prominent example, consider how digital technology facilitates the unauthorized and unpaid use of music files. Most other observers, along with me, agree that these new products—free and nearly perfect substitutes for music files
available for sale—cannibalize the revenue of music files. Similarly, my research shows that unpaid consumption of entire movies (whether copied DVDs or downloaded files) cannibalizes paid consumption of movies.

46. Contexts outside of music and movies provide less clear evidence of cannibalization. For example, the emergence of short video segments at sites such as YouTube provide an interesting contrast to the availability of free music files. It is intuitive that short segments of video programming may serve as “teasers” that stimulate interest in the full-length authorized programming. And my own research finds strong evidence that while unauthorized file sharing in music cannibalizes demand for the genuine product, unauthorized distribution of short video segments does not. It is perhaps not surprising, then, that rights holders began streaming their own authorized videos of their television programming less than a year after YouTube’s emergence in early 2005. Many programs were available late in 2005; the majority of the network schedules were available online by the start of the 2006-2007 television season.

47. A lesson emerging from my research in this area is that the free availability of a close substitute for an authorized product decreases demand, while the availability of only an element of a product (which is not a close substitute for the entire product) does

---


not depress demand. Rights holders in a variety of contexts have acted on these forces. Music rights holders allow retailers such as Amazon to post short segments of songs to aid consumers in their search processes.

48. These findings and facts have relevance to the matter at hand. The mere searchability of the text inside of a book, without any ability to read that text, is not a close substitute for access to the book. It is, instead, only a product component—akin to the short samples of songs that many music retailers make available online—and allows a potential user to know more about the work prior to using it. Similarly, many rights holders for written works allow their works to be searchable online. For example, rights holders are invited to provide Amazon a searchable version of their works, along with permission to make the search functionality available to shoppers.43 Interestingly, Amazon does not compensate rights holders for this service. Instead, rights holders receive indirect compensation in the form of sales stimulation. Amazon’s own data indicates that participation in their “Search Inside the Book” program raises sales for participating books by 9 percent.44 When Amazon launched this program in 2003, 190 publishers had already allowed their books to be searched, without remuneration, and this number has only grown since then.45

49. The sales-stimulating effects of Amazon’s “Search Inside the Book” program, along with extensive participation from rights holders, makes it clear that, rather than

cannibalizing demand, searchability stimulates demand for the books that are searchable. This finding is very much in line with my research findings that the free availability of close substitutes cannibalizes sales, while the free availability of incomplete elements stimulates demand.

50. I conclude that a service offering searchability inside books, like the HathiTrust does, would very likely benefit rights holders by stimulating demand for their works.

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

Date: June 29, 2012

Joel Waldfogel
Appendix A

Curriculum Vitae of Professor Joel Waldfogel
Joel Waldfogel
updated: January 3, 2011

Carlson School of Management
University of Minnesota
Minneapolis, MN 55455
phone: (612) 626-7128
fax: (612) 626-1335
jwaldfog “at” umn.edu
http://www.tc.umn.edu/~jwaldfog

Education
Ph. D. 1990, Economics, Stanford University
B.A. 1984, Economics, Brandeis University, *summa cum laude*, with Highest Honors

Work Experience
Carlson School of Management and Department of Economics (by courtesy), University of Minnesota
Frederick R. Kappel Chair in Applied Economics, 2010-

National Bureau of Economic Research, Cambridge, MA
Research Associate, Industrial Organization and Law and Economics Programs, 2002-present

Faculty Research Fellow, Law and Economics Program, 1993-2002

Member, Board of Advisers, Tango Card, Inc., 2010-

Member, Board of Advisers, HowMutch, 2010-

Business and Public Policy Department, The Wharton School, University of Pennsylvania
Joel S. Ehrenkranz Family Professor, 2003-2010
Department Chair, January 2006 – June 2009
Professor, 2001-2003
Associate Professor, 1997-2001

Associate Vice Dean for the Doctoral Program, The Wharton School, Fall 2000-2005

Associate Editor, *Information Economics and Policy*, July 2004-present


Member, Editorial Board, *B.E. Journals in Economic Analysis & Policy*, 2005-present

Consultant, The World Bank
  - Project on criminal conviction and labor market outcomes, 1997
  - Project on media use in African countries, 2007

Yale University Economics Department
  - Associate Professor, 1995-1997
  - Assistant Professor, 1990-1995

John M. Olin Visiting Faculty Fellow, Yale Law School, Spring 1994


**Service**

Member, NAS/STEP Committee on the Impact of Copyright Policy on Innovation in the Digital Era, 2010-2011
[http://sites.nationalacademies.org/PGA/step/copyrightpolicy/index.htm](http://sites.nationalacademies.org/PGA/step/copyrightpolicy/index.htm)

Consultant to the FCC Media Ownership Working Group, 2001-2003


Scientific Organizing Committee, Conference on Media Economics (Bologna, 2007; Zurich, 2008; Siena 2009; New York, 2010)

Co-organizer, 2009 NBER Summer Institute Conference on Industrial Organization


**Distinctions**

Journal of Urban Economics, Highly Cited Author Award, 2004-2008

Marc and Sheri Rapaport Undergraduate Core Teaching Award, The Wharton School, May 2008
Journal of Industrial Economics “Best Article of the Year” Prize, 2006

Monroe-Paine Lecture, University of Missouri, October 2008

First Prize, Mexican Law and Economics Association, for “Do Sentencing Guidelines Raise the Cost of Punishment?” (with Jose Meade), October 1998

Teacher of the Year 1994, Yale Graduate Economics Club


John M. Olin Fellow, Stanford Law School, Summer 1988

Grants

University of Pennsylvania, Wharton Global Research Initiative, 2009, 2010 (with Fernando Ferreira, to support research on globalization of recorded music industry)

University of Pennsylvania, Real Estate Center, 2004, 2005 (to support research on chain restaurants)

University of Pennsylvania, Mack Center, 2007, 2008, 2009 (to support research on video downloading and the pricing of digital products)

University of Pennsylvania, Webl/Mack Center, 2004, 2005 (to support research on music downloading and recording industry innovation)

University of Pennsylvania, Webl, 2002 (to support research on information intermediaries)

University of Pennsylvania, Wharton Electronic Commerce Forum, 2000 (to support research on the digital divide)

University of Pennsylvania Research Foundation, 1998 (with Kevin Volpp, to support research on the impacts of New Jersey health care reform)

University of Pennsylvania Research Foundation, 1998 (with Jason Scott Johnston, to support research on litigation)

Yale Social Science Faculty Research Fund, 1991, 1994 (with Steve Berry)

NSF Law and Social Science grant no. SBR-9310526, a two year grant with Ian Ayres to support research on race discrimination in Connecticut bail setting, 1993
Books


Scholarly Articles in Journals


   • Recipient of Journal of Industrial Economics “Best Article of the Year” Prize, 2006


   • Reprinted in *Problèmes Économiques*, December 22, 1999


• Reprinted in Ian Ayres, *Pervasive Prejudice?*, Univ. of Chicago Press, 2001


**Other Publications (Chapters, etc.)**


**Active Working Papers**


Inactive Working Papers


75. “A Citation-Based Test for Discrimination at Economics and Finance Journals,” (with Scott Smart), NBER working paper 5460, February 1996

76. “Bargaining in the Shadow of the Judge,” (with Orley Ashenfelter), mimeo, Princeton University, February 1993

Reports


78. “Conviction and Labor Market Outcomes: the Existing Literature and its Possible Relevance to Developing Countries,” prepared for The World Bank, October, 1997

Doctoral Students Advised
Wharton
Member, Ben Shiller
Chair, Brett Danaher, Wellesley College.
Member, David Song
Member, Leslie Schafer, PricewaterhouseCoopers.
Member, Ted Goodman, University of Arizona
Member, Gilbert Gimm, Mathematica
Member, Hart Posen, University of Michigan
Chair, Mike Gessner. Economic Analysis LLC.
Member, Ted Buckley, McKinsey
Member, Mike Furukawa, Arizona State Univ..
Member, Gus DeFranco, 2004, Univ. of Toronto, Accounting
Chair, Lisa George, 2001, Hunter College
Member, Phanish Puranam, 2001, London Business School
Member, Kate Bundorf, 2000, Stanford Medical School
Member, Kevin Volpp, 1998, Penn Medical School
Member, Linda Bornyaz, 1998

Yale
Chair, Jose Meade, 1997, ITAM
Chair, Doug Leslie, 1997, Veteran’s Administration
Member, Jessica Holmes, 1998, Middlebury College
Member, Kristin Mancini, 1997,
Member, David Popp, 1997, Syracuse University
Member, Peter Siegelman, 1991, University of Connecticut Law School
Member, Emmanuel Thorne, 1992

Recent Seminars

2010/2011
- Technology Policy Institute Aspen Forum, Aspen Colorado, August 2010
- NYU Stern Economics Group, October 2010
- Economics of Media Conference, Hunter College, October 2010
- University of Wisconsin-Madison, Department of Economics, November 2010
- FCC, December 2010
- Curb Center, Washington DC, December 2010
- Keynote speaker, WISE 2010, St Louis, December 2010

2009/2010
- University of Illinois, September 2009
- University of Texas, McCombs, October 2009
- Ohio State Economics, December 2009
- London Business School, December 2009
- London School of Economics (public talk), December 2009
- Royal Society for the Arts (public talk), London, December 2009
University of Minnesota, Carlson School, January 2010
Carnegie Mellon, Heinz College, February 2010
NBER Winter IO Meetings, Palo Alto, February 2010
Invited Speaker, Society for Economic Research on Copyright Issues (SERCI) Congress, Cartagena, Colombia, July 2010

2008/2009
LSE, June 2009
Invited Speaker, TILEC (Tilburg) conference on law and economics of media, June 2009
TPRC, Washington DC, September 2008
INFORMS, Washington DC, October 2008
QME, New York, October 2008
Ohio State University, October 2008
University of Michigan, November 2008
Kellogg, Northwestern, December 2008

2007/2008
Conference on Media Economics, Bologna, October 2007
Keynote Speaker, Economics Network for Competition and Regulation (ENCORE), May 2008, Hilversum, the Netherlands
Invited Speaker, Economics of Information and Communication Technologies, ZEW, July 2008, Mannheim, Germany
NBER Summer Institute IO, July 2008.

2005/2006
NBER Summer Institute IO (discussant)
NYU IO Day
Fundacion Ramon Areces (Madrid)
University of Chicago GSB
AEA Meetings, Boston (Presenter: Urban Economics, Internet and Entertainment Industries; Discussant: Innovation)
St. Johns Law School, Media Diversity Conference
Wharton, Mack Center Mini-Conference, Discussant

2004/2005 Seminars
Yale University Econ/SOM
USDA Product Differentiation Conference
Harvard Business School Strategy Group
Cornell Economics
Berkeley/Haas
NBER Urban Economics Conference
Jan 2005 AEA Meetings (discussant)
• European Center for Advanced Research in Economics and Statistics, Belgium
• ENST, France

2003/2004 Seminars
• NYU IO day
• Syracuse
• Dartmouth/Tuck
• Brookings Institution
• Wharton Decision Processes seminar
• Stanford Graduate School of Business (Econ & Political Economy Seminars)
• Wharton Management Department, Evolution of Organizations and Industries Seminar
• Ford Foundation Media Conference at Fordham Univ.
• Wharton Marketing Seminar
• NBER IO Winter Meetings (discussant)
• University of Toronto, Rotman School
• Kellogg
• Wharton Management Strategy and the Business Environment, discussant
• Columbia University media ownership conference
• Wharton Summer Applied Economics Seminar
• NBER Law & Econ Summer Institute (scheduled)
• NBER Innovation Summer Institute (discussant, scheduled)
• NBER IO Summer Institute (discussant, scheduled)

2002/2003 Seminars
• Speaker at University of Pennsylvania Economics Day
• Panel Member, Mediatank Public Forum on Media Ownership in Philadelphia with FCC Commissioner Michael J. Copps (May 7th, 2003)
• Wharton Summer Applied Economics Seminar
• Virginia Econ
• Carnegie Mellon University Strategy Seminar
• MIT Econ
• NBER Winter Industrial Organization Meetings, Stanford

2001/2002 Seminars
• Harvard University Econ
• George Mason University
• Rutgers University Econ
• NYU (Stern)
• University of Pennsylvania (Wharton)
• Dartmouth (Tuck and Econ)
• NBER Innovation Summer Institute

2000/2001 Seminars
• 28th Annual Telecommunications Policy Research Conference
• University of Wisconsin, Madison, Economics Department
• Stanford GSB
• Duke Fuqua
• Washington Univ (Olin)
• Michigan State Univ Econ Dept.
• Univ of Illinois Econ Dept.
• Brandeis University Economics Dept.
• Cal – Berkeley (Haas)

Other Professional Activities


Personal

Married to Mary J. Benner; two children (born 12/93 and 2/96)

University of Pennsylvania Service

2008/2009
• Member, Planning Committee, Penn Program on Democracy, Citizenship, and Constitutionalism
• Member, Wharton School Publishing Editorial Board
• Member, Wharton School Committee of Faculty Attraction and Retention
• Chair, Business and Public Policy Search Committee

2007/2008
• Chair, Marketing Department Quinquennial Review Committee
• Member, Wharton School Faculty Survey Steering Committee
• Member, Planning Committee, Penn Program on Democracy, Citizenship, and Constitutionalism

2004/2005
• Chair, Wharton Undergraduate Curriculum Review Committee
• University of Pennsylvania, Council of Graduate Faculties, member
• Wharton Doctoral Executive Committee, member
• Co-Organizer, Wharton Scholars Undergraduate Research Program

2003/2004
• Wharton Applied Economics Seminar, co-organizer
Chair, Business and Public Policy Doctoral Admissions Committee
Member, BPUB Q-Review Committee
Business and Public Policy Search Committee, Chair
University of Pennsylvania, Council of Graduate Faculties, member
Wharton Doctoral Executive Committee, member
Co-Organizer, Wharton Scholars Undergraduate Research Program

2002/2003
Wharton Personnel Committee, member
Wharton Doctoral Executive Committee, member
Wharton Applied Economics Seminar, organizer
Business and Public Policy Doctoral Coordinator
Business and Public Policy Search Committee, Chair
University of Pennsylvania, Council of Graduate Faculties, member
University of Pennsylvania, Committee on Doctoral Degree Requirements, Rules and Regulations, member

2001/2002
Wharton Personnel Committee, member
Wharton Doctoral Executive Committee, member
Wharton Applied Economics Seminar, co-organizer
Business and Public Policy Doctoral Coordinator
University of Pennsylvania, Council of Graduate Faculties, member
University of Pennsylvania, Strategic Planning Committee (Doctoral Education)

1999/2000
Wharton Research Policy Committee, member
Wharton Applied Economics Seminar, co-organizer
Public Policy and Management Ph.D. Program Committee
Wharton Ad Hoc Committee on the Doctoral Program, chair
Wharton Summer Applied Economics Seminar, founder and organizer

1998/1999
Wharton MBA Curriculum Review Committee, member
Wharton Research Policy Committee, member
Wharton Applied Economics Seminar, co-organizer
Public Policy and Management Ph.D. Program Committee
Wharton Summer Applied Economics Seminar, founder and organizer

1997/1998
Wharton MBA Curriculum Review Committee, member
Public Policy and Management Ph.D. Program Committee
Wharton Summer Applied Economics Seminar, founder and organizer
Appendix B

Cases in which I’ve provided testimony in the past four years

Distribution of the 2004 and 2005 Cable Royalty Funds (Docket No. 2007-3 CRB CD 2004-2005)
Appendix C

Materials Relied Upon

Documents Cited

Legal Documents

Deposition of Daniel Clancy, June 1, 2012.

Deposition of Frederic L. Haber, June 4, 2012.

Public Documents


Bloomberg LP.


Websites

ALA Library Fact Sheet 1, https://www.ala.org/ala/professionalresources/libfactsheets/alalibraryfactsheet01.cfm.


