LEARNING FROM GOOGLE BOOKS:

BEST PRACTICES FOR LIBRARIES SEEKING TO OUTSOURCE

THEIR DIGITIZATION ACTIVITIES

SARAH MCDONALD

This dissertation was submitted in part fulfilment of requirements

for the degree of MSc Information and Library Studies

DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES

UNIVERSITY OF STRATHCLYDE

SEPTEMBER 2017

DECLARATION

This dissertation is submitted in part fulfilment of the requirements for the degree of MSc of the University of Strathclyde.

I declare that this dissertation embodies the results of my own work and that it has been composed by me. Following normal academic conventions, I have made due acknowledgement to the work of others.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to provide copies of the dissertation, at cost, to those who may in the future request a copy of the dissertation for private study or research.

I give permission to the University of Strathclyde, Department of Computer and Information Sciences, to place a copy of the dissertation in a publicly available archive.

I declare that the word count for this dissertation (excluding title page, declaration, abstract, table of contents, and references) is 19,062.

I confirm that I wish this to be assessed as a Type 1 Dissertation.

Signature:

Date:

ABSTRACT

This dissertation investigates the contracts between Google and libraries for digitization of the libraries' collections. Research has revealed that there is scant scholarly literature concerning the contracts libraries enter into with Google as part of the Google Books Library Project. For the purposes of this paper, four contracts between Google and its library partners were located and reviewed. The goal of this research was to assemble a body of best practices for libraries that wish to digitize their collections. These best practices apply both to libraries seeking to contract with Google for digitization as well as libraries wishing to contract with another third-party, for-profit company. The discussion emphasizes the importance of libraries to be educated about the nuanced contract terms they negotiate. Once a library has decided to contract with Google or another third party for digitization, the library must consider what contract terms best suit its needs. Some things libraries should consider are the confidentiality of the contract terms, non-exclusivity provisions, time limits, the allocation of costs between the library and the digitizer, the breadth of digitization, whether the library chooses to digitize copyrighted works or whether it will limit digitization to public domain works, which party will be responsible for creating and providing metadata, whether the works will be available for data mining, how many digital copies the library will receive, the library's ability to use the digital copies on its website and for inter-library loans, the digitizer's ability to use the digital copies, both parties' rights to commercially profit from the digital copies, the library's remedies in the event of the digitizer's breach, user privacy, and the parties' responsibilities for security measures. In addition to best practices, the research also led to a compilation of alternatives for libraries that wish to digitize their collections in other ways besides partnering with Google in the Google Books project.

TABLE OF CONTENTS

I. INTRODUCTION

Google Inc. ("Google") and its products have become so standard in today's world that it is often easy to forget how controversial and innovative some of them were at their inception. One such example is the Google Books project, which was born out of Google's idealistic desire to digitize all the world's books. Almost from the project's inception, however, Google faced major hurdles in the form of litigation. Although Google ultimately prevailed, and the resulting court decision is arguably one of the most important holdings in United States copyright law in recent years, Google has fallen far short of its goals with regard to the Google Books project. Yet through these struggles and triumphs much can be learned. The court's holding, Google's practices, and Google's contracts provide valuable lessons for libraries deciding whether to engage with a third party for digitization.

A. Research Question

This dissertation investigates the contracts between Google and libraries for digitization of the libraries' collections. The research question to be answered is: Does an analysis of the contracts libraries previously entered into as part of the Google Books project offer any insights for libraries today that wish to enter into digitization contracts with Google or another third-party, for-profit company? Research has revealed that there is scant scholarly literature concerning the contracts libraries entered into with Google as part of the Google Books Library Project, and thus this is a novel research question. This is an area ripe for further exploration, as it not only combines the author's legal and library science background, but the hope is that it will also help libraries with best practices during the contract negotiation process.

B. Research Method

This dissertation is based on an extended literature review. This method is appropriate because there is an important gap in knowledge: thorough, recent legal analysis of the contracts between Google and libraries. This literature review will add to the existing body of knowledge with new analysis and best practice recommendations (Zorn and Campbell, 2006, p. 178). A literature review is the perfect vehicle for summarizing the existing literature and synthesizing it in a way that allows a new perspective (Boote and Beile, 2005, p. 4). The analysis and research can be accomplished solely via a review of the existing literature and case law, synthesizing the same, and providing new insights. Interviews, surveys, and experiments are not needed, and thus the other available research methods would be inappropriate for this topic. Further, due to the time limitations, a literature review was much more feasible than interviews or surveys. Since this dissertation hopes to add to knowledge in a field and answer specific questions dealing with libraries' contracts, the research method is more akin to a systematic literature review rather

than a narrative literature review. The main sources of information were the legal databases of Lexis Library (Law), HeinOnline, and Westlaw.

This dissertation begins with background information about digitization, the Google Books website and Library Project, and the resulting Google Books lawsuit. The dissertation then moves into a discussion of the criticisms and benefits of the Google Books project. Following this is an in-depth analysis of some of the contracts between Google and its library partners, highlighting key similarities and differences between the publicly-available contracts. The heart of the dissertation is a set of best practices created from the review of the Google contracts. Included are suggestions about what libraries should consider when deciding to enter into a digitization contract with Google or another for-profit company. Finally, the dissertation concludes with several digitization options available to libraries that do not wish to partner with Google.

Conducting a literature review and performing a legal analysis of the contracts between Google and libraries have led to a deeper understanding of the advantages and disadvantages of partnering with Google in a library context. The hope is that this analysis – and especially the recommendations about best practices – will prove useful to scholars or librarians facing the question of whether to partner with Google or another for-profit company for digitization. Finally, this analysis fills in a gap that currently exists in the literature about this subject.

II. LITERATURE REVIEW

A. Digitization

The Google Books project is, of course, about books. But the project is not just about any book – specifically, it is about e-books. Before delving into any legal issues surrounding e-books and digitization, a definition and brief history are needed.

The Oxford English Dictionary (2001) defines "e-book" as "a hand-held electronic device on which the text of a book can be read." Less common definitions are "a book whose text is available in an electronic format for reading on such a device or on a computer screen" and "a book whose text is available only or primarily on the Internet" (Oxford English Dictionary, 2001). The e-books available through the Google Books project fit the former, less common definition, as they are digitized versions of existing books. Practically speaking, most people think of an e-book as something they purchase online or borrow from their local library and read on a portable device such as their phone, computer, tablet, or e-reader. With the Google Books project, Google hopes to expand this common perception to include full-text versions of out-of-copyright works and partial versions of in-copyright works available through its website.

There are many types of e-book formats: printable PDFs ("portable document format"), enhanced PDFs with hyperlinks and embedded files, and "true e-book" formats such as EPUB or MOBI files (Bartram, 2014). EPUB is the most common and non-proprietary format, used by most United States government agencies, commercial publishers, and libraries, and can be used on a variety of devices and software from different vendors (Bartram, 2014). MOBI or AZW is used by Amazon for its Kindle readers and software (Bartram, 2014). EPUB and MOBI are "true e-book" formats because they offer resizable and reflowable text which automatically adjusts to the font, screen size, and orientation on the device or software (Bartram, 2014).

Although the concept of e-books started much earlier (with some visionaries during the 1930s), Michael Hart was the first person to actually create the world's first e-book in 1971 (Bartram, 2014). Hart launched Project Gutenberg and digitized the United States Declaration of Independence, followed by the Bill of Rights (Manley and Holley, 2012, p. 296). Notably, Project Gutenberg remains a robust source of free e-books to this day.

There were various early experiments with e-books in the 1980s and early 1990s, but 1998 was a watershed year for a variety of reasons. The first dedicated e-book readers were launched, called Rocket E-book and Softbook (Bartram, 2014). That year also saw the first International Standard Book Number issued to an e-book (Bartram, 2014). The United States Congress passed the Digital Millennium Copyright Act ("DMCA"), which implemented major changes to copyright law to address new issues brought about by the internet, including laws that discouraged copying copyrighted materials and laws to prevent circumvention (Manley and Holley, 2012, p. 300). On the heels of the DMCA came digital rights management ("DRM"), which encompasses any mechanism, either hardware or software, that attempts to control the use and distribution of intellectual property in digital form (Manley and Holley, 2012, p. 300). Types of DRM systems include data encryption, digital watermarks, and user plug-ins (Manley and Holley, 2012, p. 300). E-books were especially impacted because most DRM controls either prohibit or limit copying and often do not allow an e-book to be transferred to a new device when coupled with a restrictive licensing agreement (Manley and Holley, 2012, p. 300). As a result, companies like Amazon and Apple developed proprietary formats that limited users to their e-book inventory and made it impossible to sell e-books through the doctrine of first sale because the e-books are licensed rather than purchased (Manley and Holley, 2012, p. 300). DRM is particularly challenging to libraries who purchase materials intended for multiple users (Manley and Holley, 2012, p. 300). Also in 1998, libraries in the United States began providing free e-books to the public through their websites and associated services (Bartram, 2014). Finally, it was in 1998 that Larry Page and Sergey Brin founded Google (Bartram, 2014); Google Books would follow shortly thereafter.

Fast forward to 2007, and the world of e-books was drastically changed with Amazon's introduction of the first Kindle and Apple's introduction of the first iPhone (Bartram, 2014). At about the time the Kindle was introduced, a little more than one percent in total publishing trade revenue went to e-book sales; just four years later, that had jumped to 23 percent, attributed in large part to the Kindle (Inouye, 2016, p. 42).

In 2009, Barnes & Noble introduced the Nook e-reader and Sony linked with libraries via the OverDrive digital network to enable library patrons to borrow e-books from their local libraries (Bartram, 2014). The following year ushered in other new advancements for e-books. Apple released the iPad and put its iBooks and iBookstore on iTunes, selling half a million e-books in less than a month (Bartram, 2014). That year also saw the launching of Google's e-bookstore (Bartram, 2014), not to be confused with Google Books. Amazon reported that its e-book sales outnumbered its hardcover book sales for the first time, and shortly thereafter, in January 2011, Amazon reported that its e-book sales outnumbered its paperback sales (Bartram, 2014).

By 2011, Amazon owned two-thirds of the e-book market, and in April of that year it announced that Kindle readers in the United States would be able to borrow library e-books by December (Woo and Trachtenberg, 2011). Until then, library users who borrowed e-books could read them on Barnes & Noble's Nook, the Sony Reader, the Kobo reader, and on laptops and smartphones (Bosman, 2011). Some American librarians believe that Amazon's decision proves that libraries are a key player in the e-book business and that Amazon was unable to overlook them any longer (Bosman, 2011).

The biggest complaint from library patrons about e-books is the lack of availability from their local libraries (Milliot, 2015). In reality, libraries face many hurdles in the procurement of e-books, and unfortunately the library patrons are the ones who suffer. As recently as a few years ago, major publishers made it very difficult – and in some cases impossible – for libraries to purchase and lend e-books (Vaccaro, 2014). Although the situation has improved slightly, publishers still put restrictions on e-books bought and offered by libraries. Some publishers only allow for an e-book to be borrowed 26 times before the library has to purchase the license again, others opt for the license to expire after a year, others charge libraries significantly more for e-books than they charge consumers, and others charge libraries significantly more for e-books than print books (Vaccaro, 2014). Such restrictions would violate copyright law for print books, but because e-books are different (libraries purchase a license rather than a digital copy), publishers can engage in these tactics (Vaccaro, 2014). Some libraries have fought back against what they see as unfair treatment by publishers. The American Library Association also fought back, and created the Digital Content & Libraries Working Group in late 2011 to address concerns that

publishers were restricting or refusing access to e-books through libraries (American Library Association, 2017). The group meets with the nation's largest publishers as well as distributors, author groups, and other representatives, in an effort to ensure that libraries are not taken advantage of when it comes to procuring e-books (American Library Association, 2017).

The British Library quickly embraced e-book technology and by the end of 2011, it offered Kindle owners the ability to view first editions from such notable British authors as Charles Dickens and Jane Austen in the original typefaces, with original illustrations (Hall, 2011). In addition, at that time the entire collection digitized by Microsoft became available to view on Apple's iPad (Hall, 2011).

Despite the many advantages of e-books and the public's general awareness of their existence at local libraries, a low usage rate plagues libraries in the United States. More than 90 percent of public libraries offer e-books to their patrons, but only 62 percent of adults are aware of these offerings, according to "Libraries and Learning," a 2016 report from Pew Research Center's Internet, Science & Tech division (Arch, 2016, p. 1). A 2015 study found that even though over half of library patrons surveyed were aware that their local libraries carry e-books, only 25 percent had actually borrowed an e-book within the past year (Milliot, 2015). Patrons' biggest complaint was the lack of e-books' availability, followed by a preference for print books and dissatisfaction with the loan period for e-books (Milliot, 2015).

These statistics imply that libraries may need to do a better job to get the word out about e-books. Yet libraries face some real challenges with the promotion of e-books, including lack of time on the part of libraries to promote them and train patrons how to use them, lack of availability both in the market and in the libraries' collection, budget constraints, the public's lack of interest, and the hurdle of promoting a technology that many still view as new (Vasileiou and Rowley, 2011, p. 634). In addition, librarians justifiably have concerns about not raising expectations about access to e-books, because publishers have a significant role in determining the availability of, and licensing arrangements associated with, e-books (Vasileiou and Rowley, 2011, p. 638). No library wants to be put in a position of promoting e-books that end up being unavailable, either because the publisher has refused to sell the title to the library, the library does not have the budget to purchase the title, or there is a long wait for the library's single copy of the title.

An interesting recent phenomenon that is being blamed for the stagnant or falling rates of e-reading is "digital fatigue" (Milliot, 2016). Somewhat surprisingly, younger people are expressing a desire to disengage from their screens at a much higher rate than older people (Milliot, 2016). Since consumers almost always have the option to read books in print form, they are indicating a preference to return to that (Milliot, 2016). In fact, e-book sales declined and print book sales actually increased in 2015 and

2016, prompting one commentator to declare that publishers are "back to good old paper" (Bershidsky, 2017). Reading a paper book "is a statement, a human being's answer to being increasingly surrounded, and now even threatened, by machines" (Bershidsky, 2017).

B. Google Books Website

In essence, the Google Books website is a Google search engine for books. Users can search for books via a standard Google search, author, title, International Standard Book Number, Library of Congress Control Numbers, and Online Computer Library Center record numbers (Google Inc., 2017a). The website displays book results one of four ways depending on the books' copyright status:

> Full view: if a book is out of copyright or the publisher or author has asked to make the book fully viewable, the user can view any page from the book and, if the book is in the public domain, download, save, and print a PDF version.
>
> Limited preview: if the publisher or author has given permission, the user can see a limited number of pages from the book.
>
> Snippet view: if the book is in copyright, the user can view information about the book plus a few snippets (which are a few sentences to display the user's search term in context).
>
> No preview: the user can only see bibliographic information ("like a card catalog") (Google Inc., 2017a). Since 2005, Google will exclude any book from snippet view at the copyright holder's request and via the submission of an online form (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210).

Much was made in the lawsuit regarding the "snippet" view. A snippet is a horizontal segment of text usually comprising one-eighth of a page (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209). Each page of a conventionally formatted book in the Google Books database is divided into eight non-overlapping horizontal segments, and thus each horizontal segment is a snippet (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209-210). If a search term appears three or more times within a particular book, the same three snippets will appear on the Google Books website after searching for that term, regardless of the number of users who search for that term or the number of computers used to perform the search (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). Further, even if a particular term appears many times within a page, the Google Books website only displays the first usage of that search term (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). Thus, if the top snippet of a page contains two (or more) words for which the user searches, and Google's program is fixed to reveal that

particular snippet in response to a search for either term, the second search will duplicate the snippet already revealed by the first search, rather than moving to reveal a different snippet containing the word because the first snippet was already revealed (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). The Google Books algorithms do not allow circumvention of the snippet view by users. To that end, if a user performs multiple searches of the same term from multiple computers, the Google Books website will only reveal the same snippets (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). The only way a user can view more than three snippets of a book is by entering additional searches for different terms (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). However, Google employs what it calls "blacklisting" by making certain portions of the book unavailable regardless of whether those portions contain search terms (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). The Google Books website permanently blocks one snippet on each page and one complete page out of every ten (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). Google also disables snippet view entirely for certain types of books such as dictionaries, cookbooks, and books of short poems (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). The understanding is that for these types of books, it is likely that a single snippet would satisfy the user's present need for the book (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210).

All books on the Google Books website have links to relevant bookstores and Online Computer Library Center WorldCat (Google Inc., 2017a). This feature allows users to find books for purchase or at their local libraries.

C. Library Project

Google obtains books for the Google Books website via one of two ways: the Partner Program, in which Google partners with publishers, and the Library Project, in which Google partners with libraries (Google Inc., 2017a). The Google Books Library Project began in 2004 with the libraries of Harvard University, Stanford University, the University of Michigan, the University of Oxford, and The New York Public Library (Google Inc., 2004). What began with just five libraries has expanded greatly – Google now has contracts with over 40 libraries around the world to digitize their collections (Google Inc., 2017b).

Under the contracts, participating libraries select books from their collections to submit to Google for inclusion in the project (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 208). Google creates a digital scan of each book, extracts machine-readable text, and creates an index of the machine-readable text of each book (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 208). Google retains the original scanned image of each book, in part to improve the accuracy of the machine-readable texts and indices as

image-to-text conversion technologies improve (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 208).

Google made no secret about the fact that it was actively seeking out libraries to partner with for the Google Books project, even more so than publishers. It may be that Google started wooing libraries more than publishers to contribute books to the Google Books project simply because of the sheer volume of books libraries could offer – notably out-of-copyright books. Within the first few years of the Google Books project, a handful of library partners were able to offer Google access to 15 to 20 million books (Newman, 2011). This is in stark contrast to the thousands of publishers in the Partner Program, each with perhaps only a few hundred titles (Newman, 2011).

Google has not released a comprehensive list of the 40-plus libraries it has contracted with, choosing instead to highlight 21of these libraries on its "Library Partners" website (Google Inc., 2017b). These highlights, however, are silent as to the date the particular library joined the project or information about the library's contract with Google. Piecing together information from various sources, a timeline of nearly all of Google's library partners is as follows:

> December 2004: the project begins with the libraries of Harvard University, Stanford University, the University of Michigan, the University of Oxford, and The New York Public Library (Google Inc., 2004).
>
> August 2006: the University of California System joins the project, with an agreement to digitize 2.5 million books (Carlson, 2006).
>
> September 2006: the Complutense University of Madrid, which houses Spain's largest university library, becomes the first non-English language library to join the project (Woo, 2006). To avoid any copyright issues, the university will allow only works in the public domain to be scanned into Google's database (Woo, 2006).
>
> October 2006: the University of Wisconsin joins the project along with the Wisconsin Historical Society Library (University of Wisconsin, 2006). Combined, the libraries contain more than 7.2 million holdings (University of Wisconsin, 2006).
>
> November 2006: the University of Virginia Library joins the project (Wood, 2006). Built by Thomas Jefferson, one of the founding fathers of the United States, the library contains a vast collection of early American historical material among its five million volumes and 17 million manuscripts, rare books, and archives (Wood, 2006).

January 2007: the University of Texas at Austin joins the project (The University of Texas at Austin, 2007). The multi-year partnership anticipates that Google will digitize at least one million volumes from the libraries' collections, working from selection lists prepared by the university (The University of Texas at Austin, 2007).

January 2007: the National Library of Barcelona, Catalonia's largest library, and four affiliate Catalonian libraries join the project to digitize their public domain works in an effort to further their mission of collecting, preserving, and spreading Catalonian bibliographic production and that related to the Catalonian linguistic area (Campon, 2007).

February 2007: Princeton Library joins the project and agrees to contribute one million public domain books (Cliatt, 2007).

March 2007: the Bavarian State Library (Bayerische Staatsbibliothek) in Munich, Germany, becomes the largest non-English language library to join the project and will contribute more than one million out-of-copyright books in German, English, French, Spanish, Italian, and Latin (M2 Best Books, 2007).

May 2007: Ghent University Library becomes the first Dutch language library to join the project (Google Inc., 2011).

May 2007: the Cantonal and University Library of Lausanne, Switzerland, becomes the first French language library to join the project (Bangeman, 2007).

June 2007: the 12 universities of the Committee on Institutional Cooperation (now known as the Big Ten Academic Alliance) join the project to digitize 10 million volumes in their collections (Albanese, 2007). The member universities are the University of Michigan and the University of Wisconsin (which had already signed contracts with Google), the University of Minnesota, the University of Chicago, the University of Illinois, Indiana University, the University of Iowa, Michigan State University, Northwestern University, Ohio State University, Pennsylvania State University, and Purdue University (Albanese, 2007).

July 2007: Keio University becomes the first Asian library in the project and plans to digitize at least 120,000 books (Jiji Press English News Service, 2007). Due to copyright protection, digitization will be limited to academic books for which the university holds the copyright or books whose copyright has expired (Jiji Press English News Service, 2007).

August 2007: Cornell University Library becomes the 27th partner and Google will digitize up to 500,000 copyrighted and public domain items from its collection (Research Information, 2007). The university announced that Google will provide a digital copy of all the materials scanned, which will eventually be incorporated into the university's own digital library (Research Information, 2007).

December 2007: Columbia University Libraries, one of the top 10 academic library systems in the United States, joins the project and agrees to contribute public domain works from its 25 distinct libraries (Columbia University Libraries, 2007).

July 2008: the Municipal Library of Lyon in France's second largest city becomes the first French library to join the project and agrees to contribute 500,000 public domain books (Colombet, 2008).

March 2010: the National Libraries of Florence and Rome, through the Italian Ministry of Cultural Heritage, will contribute up to a million out-of-copyright works (Mattiuzzo, 2010). These are the first Italian libraries and the first ministry of culture to join the project (Mattiuzzo, 2010).

June 2010: the Austrian National Library becomes the tenth library partner in Europe and will contribute up to 400,000 public domain books from its collection (Weisl, 2010). The books will be available for inclusion in Europeana and other digital libraries (Weisl, 2010).

July 2010: the National Library of the Netherlands (Koninklijke Bibliotheek) joins the project and will provide 160,000 volumes, almost the library's entire collection of out-of-copyright books, that were written during the 18th and 19th centuries (Colombet, 2010). The library will receive copies of the scans so that they can also be viewed via the library's website, and the library plans to make the digitized works available via Europeana, a European-based website which brings together the collections of various galleries, libraries, archives, and museums (Colombet, 2010).

June 2011: the British Library partners with Google Ireland Limited for up to 250,000 public domain works (British Library, 2011). The project will digitize a range of printed books, pamphlets, and periodicals dated 1700 to 1870 in a variety of major European languages (British Library, 2011). The project will focus on books that are not yet freely available in digital form (British Library, 2011).

D. Google Books Lawsuit

The Google Books lawsuit, *The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202 (2nd Cir., 2015), *cert. denied* 136 S. Ct. 1658 (2016), is arguably the most important decision involving United States copyright law in the past 25 years. The plaintiffs consisted of The Authors Guild, Inc. and individual authors. The Authors Guild is the United States' oldest and largest professional organization for writers, whose members include novelists, historians, journalists, and poets – both traditionally and independently published – along with literary agents and representatives of writers' estates (The Authors Guild, Inc., n.d.). The Authors Guild advocates for authors on issues of copyright, fair contracts, free speech, and tax fairness, and has initiated lawsuits in defense of authors' rights, such as the Google Books litigation (The Authors Guild, Inc., n.d.). The plaintiffs in the Google Books case filed their lawsuit in 2005 and asserted that the Google Books project violated United States copyright law (Young, 2005). According to the fourth amended complaint, Google contracted with several public and university libraries to create digital copies of the libraries' books (The Authors Guild, Inc., 2011, p. 2). The operative complaint further alleged that as part of the consideration for creating the digital copies, the contracts entitled Google to reproduce and retain for its own commercial use digital copies of the works (The Authors Guild, Inc., 2011, p. 2). The plaintiffs asserted that Google was committing copyright infringement by the following acts: (1) reproduction and distribution of digital copies of copyrighted works to libraries, (2) reproduction of digital copies for itself, and (3) publicly displaying the digital works in order to, among other things, attract visitors to its website and generate advertising revenue (The Authors Guild, Inc., 2011, p. 2). Judge Denny Chin of the United States District Court for the Southern District of New York dismissed the case on November 14, 2013, holding that Google Books provided "significant public benefits" and that Google's scanning of 20 million books constituted "fair use" under United States copyright law (Chant, 2013). The Second Circuit Court of Appeals unanimously affirmed the judgment on October 15, 2015 (Bloom and Collins, 2015, p. 1). The United States Supreme Court denied review on April 18, 2016, making the decision of the Second Circuit Court of Appeals final (Chandler, 2016).

At the heart of the case is the essence of copyright: that is, the right to copy. "Whether you sit down and copy a text with a quill and parchment or electronically scan and store a text using state of the art digital technology, a copy is still a copy" (Laing, 2006, p. 134). The Second Circuit Court of Appeals began its opinion by stating that "this copyright dispute tests the boundaries of fair use" (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 206). The "fair use" doctrine, as codified in Section 107 of the United States Copyright Act of 1976, states that fair use of a copyrighted work for purposes such as criticism, comment, news reporting, teaching, scholarship, or research is not an infringement of copyright (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 212). In determining whether use of a copyrighted work is fair use

under the statute, the court shall consider the following factors: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for non-profit educational purposes, (2) the nature of the copyrighted work, (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole, and (4) the effect of the use upon the potential market for or value of the copyrighted work (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 212-213).

It should be noted that the American version of "fair use" varies greatly from the British version of "fair dealing." The British concept of fair dealing provides a defense to copyright infringement, but its extent is limited and only applies to non-commercial research and private study, criticism or review, reporting current events, judicial proceedings, and certain educational purposes (Laing, 2006, p. 134). Fair dealing is a narrow defense in both scope and application and would likely not afford a defense for the broad searches made possible by the Google Books website (Laing, 2006, p. 134). The British concept of fair dealing is thus distinguished from the American concept of fair use which was at issue in the Google Books lawsuit. This point became important in the Google Books project, as all of Google's non-American partners contracted to digitize only public domain works.

As to the first factor in determining fair use, the Second Circuit concluded that Google's use of the books is transformative because it provides something that the original books do not. In fact, the purpose of Google Books is to provide users with information that is otherwise unavailable about the original texts (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 217). Google's full-text copying of the entire book along with providing snippets around search terms further this purpose (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 217). Google Books allows new forms of research known as "text mining" and "data mining" (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209). Google's "ngrams" research tool provides statistical information about the historical frequency of word and phrase usage (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209). This tool enables users to detect fluctuations of interest in a particular subject over time and space by showing increases and decreases in the frequency of reference and usage in different periods and different linguistic regions (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209). This tool also allows users to search through the millions of books Google has scanned in order to examine word frequencies, syntactic patterns, and thematic markers and to derive information on how nomenclature, linguistic usage, and literary style have changed over time (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209). One example the court gave was the way the phrase "United States" has changed over time from plural ("the United States are") to singular ("the United States is"), a topic of interest to some historians (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209). A researcher would not be able to glean this information from the print copies of the books without going through the arduous – and perhaps impossible – task of reading through the entire books

and manually recording the instances of the desired information (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209).

The plaintiffs had argued that Google was engaged in copyright infringement in part due to the amount of the work that Google copied – namely, the entire book. The court disagreed and held that this was essential to allow full-text searching of the books (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 221). The court found that not only is the copying of the entire original work reasonably appropriate to Google's transformative purpose, it is absolutely necessary to achieve that purpose (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 221). If Google copied less than the entire original, its search function could not accurately discern whether or how often a search term appears in a book (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 221).

In its discussion of the fourth factor, the court recognized that the snippet function can cause *some* loss of sales, either through users not purchasing a title or a reduction of demand on libraries for a title, which might have resulted in libraries purchasing additional copies (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 224). Yet the court found that the possibility of some loss of sales does not suffice to make the copy an effectively competing substitute (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 224). If the court had found that Google's snippet function acted as a competing substitute for the original book, such a holding would tilt the fourth factor in favor of the original work's copyright holder (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 224).

The court also dismissed the plaintiffs' arguments that Google is motivated by commercial profit and thus its use of the books cannot constitute fair use. First, the court refuted this argument by finding that there are several components of the Google Books website that do not result in profit for Google. For example, the Google Books website does not display advertising to users of the search function nor does Google receive payment when a user purchases a book through Google's link to the bookseller (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 209). Second, the court found no reason why Google's overall profit motive should be a reason to disregard a finding of fair use. Google demonstrated a "highly convincing transformative purpose," along with an absence of significant substitutive competition, and the court found these to be reasons for granting fair use (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 219). The court noted that many of the most universally accepted forms of fair use, such as news reporting and commentary, quotations, book reviews, performances, and parody, are all normally done commercially for profit (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 219).

E. Google Books Problems

In contrast to what some find as Google's blatant self-promotion, there is a plethora of literature criticizing the Google Books project. One popular topic concerns errors in the physical scanning of the books, prompting one scholar to conclude that Google has pursued a "good enough" scanning strategy in its frenzy to digitize libraries' collections (Musto, 2009). Pages are skewed, blurred, folded back, misplaced, upside down, and missing; some images contain Google employees' hands and fingers caught in the scanning process (Musto, 2009). A "small but thriving subculture" is documenting the errors in Google's scanning process, in the form of Tumblrs, printed books, photographs, online videos, and gallery-based exhibitions (Goldsmith, 2013). The most thorough documentation is a Tumblr called *The Art of Google Books*, which collects images of a printed book's history (like library due date stamps, pressed flowers, and doodles) and digital glitches that result from the scanning process (like the aforementioned hand prints) (Goldsmith, 2013). Because there has been so much written about it – and so much fascination over the hand prints in particular – one could easily get the impression that such errors are rampant throughout Google Books. This conclusion is open to interpretation. One preliminary study in 2009 determined that only one percent of tested works had errors serious enough to render the page illegible (James, 2010, p. 226). However, to a scholar, even an error rate of one percent is unacceptable, especially if the illegible page is critical for research (James, 2010, p. 227). Further, the 2009 study only analyzed significant errors; no doubt many more than one percent of books have scanning errors that do not affect the legibility of the text but are annoying and unprofessional nonetheless.

The second major problem with Google Books is the vast amount of inaccurate metadata for the books. Metadata are data about a text or work, such as title, author, publisher, date of publication, and the number of pages (Miller, 2010) – things typically found in a library catalog. Although Google has stated that it wants Google Books to be the world's catalog, the fact remains that the books' metadata are not needed to perform Google Books searches. Rather, a traditional search on Google Books is like a regular Google search – it searches the books' text. Since metadata are not essential for its version of searching, this may have led Google to not initially focus on the importance of accurate metadata when it began its book scanning process (Miller, 2010). In fact Google may be relying on full-text searching of keywords to compensate for faulty metadata (James and Weiss, 2012, p. 21). The metadata errors run the gambit and some are quite amusing. Such errors include incorrect dates (a search for books published before 1950 containing the word "internet" brings up 527 results), misattributed authors (Sigmund Freud is listed as a co-author of a book on the Mosaic Web browser and Henry James is credited with writing "Madame Bovary"), and subject misclassifications (an edition of "Moby Dick" is categorized under "Computers"

and "Jane Eyre" as "Antiques and Collectibles") among many, many others (Miller, 2010). A catalog of copyright entries from the Library of Congress is listed under "Drama," prompting one scholar to wonder if this was a joke on Google's part (Nunberg, 2009), in light of its ongoing litigation over copyright. Shockingly, Google has admitted to *millions* of errors with metadata (Reisz, 2009), and one study estimated an error rate of over 36 percent, pointing "toward a general disregard for the importance of metadata" (James and Weiss, 2012, pp. 19-20). Google blamed the errors largely on outside contractors who manually input the metadata (Miller, 2010). However, if Google Books really does want to become the world's virtual card catalog, it will need to vastly improve its metadata so that it can compete with existing library systems (James and Weiss, 2012, p. 20). Indeed, one wonders how Google Books can even be compared to a library catalog if it contains so many errors.

Although the Library Partners have no control over Google's scanning techniques and creation of metadata, some commentators have asserted that the university libraries that rushed to partner with Google in the early days of the Library Project were not particularly attentive to questions of quality (Nunberg, 2009). Indeed, the libraries' contracts with Google are silent as to remedies and procedures for these types of errors.

Such errors may lead to the conclusion that digitization is moving faster than the current technology. Google desires to digitize "every scrap of information" – which begs the question of whether Google is throwing content on the "information pile" faster than the current technology can extract the information in any useful or meaningful way (Townsend, 2007). One commentator pondered "the costs to history if the real libraries take error-filled digital versions of particular books and bury the originals in a dark archive or the dumpster" (Townsend, 2007). This is especially troublesome because of the likely high cost of fixing the errors. It is doubtful that Google will spend the money and labor to return to digitized books and correct errors in scanning and metadata.

Another serious concern about Google Books involves user privacy. The Google Books website is governed by the same privacy policy as the Google website (Google Inc., 2016, p. 4). During searches, Google collects vast amounts of information about users – whether logged into Google or not – such as device information, log information, location information, unique application numbers, and cookies (Google Inc., 2016, pp. 1-2). Google utilizes this information to offer tailored content such as more relevant search results and advertisements to its users (Google Inc., 2016, p. 2). Google reserves the right to share personal information with companies, organizations, or individuals outside of Google if Google has a good-faith belief that access, use, preservation, or disclosure of the information is reasonably necessary to: (1) meet any applicable law, regulation, legal process, or enforceable governmental request,

(2) enforce applicable terms of service, including investigation of potential violations, (3) detect, prevent, or otherwise address fraud, security, or technical issues, and (4) protect against harm to the rights, property, or safety of Google, Google's users, or the public as required or permitted by law (Google Inc., 2016, p. 4). Google also reserves the right to share non-personally identifiable information publicly and with Google's partners – like publishers, advertisers, or connected sites – to, for example, show trends about the general use of Google services (Google Inc., 2016, p. 4).

The American Civil Liberties Union (2009) does not believe that Google's Privacy Policy adequately protects reading records from disclosure to the government and third parties. Readers should be able to use Google Books without worrying that the government or a third party is "reading over their shoulder" (American Civil Liberties Union, 2009). The American Civil Liberties Union (2009) cautions against disclosing reading records without a properly-issued warrant from law enforcement and court orders from third parties.

The issue of Google Books' privacy is especially relevant for libraries, because Google and libraries have different standards for handling user information. Many libraries routinely delete borrower information, and organizations such as the American Library Association have fought hard to preserve patrons' privacy in the face of laws such as the United States Patriot Act (McMillan, 2009). "The confidentiality of patron's library use is taken very seriously by public libraries" (Madison Public Library, 2016). Public libraries must analyze their requirements to provide patrons' information, such as borrowing data and internet search histories, in response to a subpoena or warrant or as part of a Freedom of Information request. Fundamentally, this concerns patrons' rights to privacy and the ability of the government or law enforcement to override such rights. As stated in Article 8 of the European Convention on Human Rights, a public authority shall not interfere with the exercise of the right to privacy "except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic wellbeing of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others" (Council of Europe, 1988).

In the United Kingdom, police and other security organizations have the power to demand patrons' personal data under various statutes (Chartered Institute of Library and Information Professionals, 2011). If a public library receives such a request, it should ensure that the police are requesting the data for a specific purpose and are not simply on a "fishing expedition," and that proper procedures are followed (Chartered Institute of Library and Information Professionals, 2011, p. 20). The Chartered Institute of

Library and Information Professionals (2011, p. 20) is quick to point out that "[u]nless the police have invoked specific powers under one of the Acts there is no general obligation to answer police questions."

In the United States, the Patriot Act has granted law enforcement broad powers of surveillance and seizure at public libraries. "As a result of the USA Patriot Act, public libraries face a dilemma of having the responsibility of protecting the privacy of library users while responding to legitimate national security concerns" (Madison Public Library, 2016). As a result of the Patriot Act, state and local law enforcement agencies have the right to work with federal agents and apply for extended wiretapping authority, including installing monitoring devices on public library computers (American Library Association, 2005).

The reach of the Patriot Act is highlighted by the case of Joan Airoldi, a librarian in Whatcom County Rural Library District in Washington, who refused to turn over to the Federal Bureau of Investigation ("FBI") the names of every patron who had borrowed the biography "Bin Laden: The Man who Declared War on America" (Brodeur, 2005). Airoldi and the library board fought the FBI's subpoena in court, which the FBI eventually withdrew (Brodeur, 2005). Of particular concern is that if the FBI had instead presented the library with a Patriot Act order, rather than a grand jury subpoena, the library would have been unable to challenge the demand in court (Brodeur, 2005). Notably, Airoldi subsequently received the PEN/Newman's Own First Amendment Award for her actions (Prose, 2016).

It should be noted that there are situations in which a public library *must* report certain information to law enforcement. "There are some instances when failure to report a matter to the police can be a criminal offence in itself" (Chartered Institute of Library and Information Professionals, 2011, p. 20). For example, in the United Kingdom it is a criminal offense "for a person to fail to disclose, without reasonable excuse, any information which they know or believe might help prevent another person carrying out an act of terrorism or might help in bringing a terrorist to justice" (Chartered Institute of Library and Information Professionals, 2011, p. 20).

An interesting result of public libraries' own experience with government surveillance is that some libraries have begun to conduct workshops designed to teach patrons how to protect their online privacy (Prose, 2016). The first such library in the United States was Watertown Free Public Library in Massachusetts, whose librarian Alison Macrina began teaching patron computer classes about online privacy and organized a series of workshops for Massachusetts librarians about digital surveillance (Macrina and Glaser, 2014). It is unknown if Google will fight so hard to protect users' privacy.

Google Books searches are just like regular Google internet searches (Google Inc., 2017a). Therefore, book search results are generated using Google's search algorithms (Newman, 2011), and are affected by the user information that Google collects, just like regular search results (Google Inc., 2016, p. 2). Unlike a library catalog, which provides search results without regard to the user's identity or search history, Google Books displays weighted results tailored to the user. This leads to the issue of whether a user is able to obtain unbiased search results on Google Books, and, indeed, whether Google even intends such a thing to happen.

Beyond the issue of privacy, there is the issue of censorship. Books have a long history of being banned, censored, filtered, and criticized, in the general public and also within libraries. Google has previously bowed to government pressure to filter or otherwise control content on the internet (Newman, 2011). One wonders what would happen if Google faced governmental pressure to remove a book from the Google Books website.

A further issue is user access. Google Books, like Google itself, is based on the premise of universal accessibility. Indeed, Google proclaims that the Library Project makes it possible for users to search through millions of books written in many different languages, including books that are rare, out of print, or generally unavailable outside of the library system (Google Inc., 2017a), all from the comfort of their home computers. Yet this vision fails to acknowledge the varying levels of access to ready and reliable internet infrastructure (Hoffmann and Bloom, 2016). Significant disparities in internet availability, cost, and speed, as well as low or nonexistent digital literacies and skills, lead some scholars to conclude that Google Books' goal of "transcending social class and geography" is "distant and misguided" (Hoffmann and Bloom, 2016). Those who may be most able to benefit from digitized collections often lack the high-bandwidth internet connections capable of downloading and viewing high-resolution color images (Leetaru, 2008). The counterargument is that by offering books online, Google Books is opening up library collections to far more people than would be able to access them if they had to physically travel to the library to view the books. Although this may be true, one cannot ignore the reality that many people still do not even have access to online materials, let alone the ability to travel to a physical library.

F. Google Books Benefits

With so much press given to the problems with Google Books, it may be puzzling why any library would agree to partner with Google at all. The reality is that the lure of Google likely has a lot to do with cost. Libraries see a partnership with Google as being a low cost – or free – way to digitize their collections. For example, the British Library's contract with Google provides that Google will cover the full cost of digitization while the British Library will pay the lower costs of curating and caring for the selected books

(Allsop, 2011). In fact, although Google has not disclosed how much it is spending on the British Library project, it spent $200 million on digitization projects at four libraries in the United States (Allsop, 2011). Google has not been forthcoming with the amount of money the Google Books project has cost per library. However, even if such costs amounted to a fraction of the $200 million referenced above – say, $10 million – that is still likely way out of the league of most libraries, and certainly beyond the reach of small or nonacademic libraries. Google has an army of human page-turners, the ability to leverage its existing server farms, and reliance on an almost entirely automated backend processing environment – which result in Google's ability to scan books at a cost of just $10 each (Leetaru, 2008). The average library does not have anywhere near this capability.

The Online Computer Library Center found that most libraries, regardless of size, overwhelmingly identified insufficient staff time, insufficient ongoing funding, and inadequate technology and equipment as major barriers to digitizing their collections (Morgan and Proffitt, 2017, p. 11). Digitization is careful, time-consuming, and expensive work. Scanning old, fragile, one-of-a-kind works has its own obstacles, and "sometimes resulted in the death of a book" (Clements, 2009). The software commonly used to scan books, Optical Character Recognition ("OCR") software, requires each page of the book to be flat (Clements, 2009). One way to flatten a book is to use glass plates that individually flatten each page; another way is to remove the book's binding, thus destroying the book in the process (Clements, 2009). Google eliminated the need for these tedious and destructive methods by inventing infrared camera technology that detects the three-dimensional shape and angle of book pages when the book is placed in the scanner (Clements, 2009). This information is transmitted to the OCR software, which adjusts for the distortions and allows the OCR software to read text more accurately (Clements, 2009).

Outside of the Google Books arena, digitization in general offers many benefits to libraries. The first is preservation. In fact, digitization as a method of preservation is so cost effective and has such a positive impact on organic materials like paper that it is "exceedingly difficult to defend the allocation of financial resources for physical conservation treatments, preservation rebinding, special collections rehousing, and almost every other preservation action that involves the handling of individual physical artifacts" (Conway, 2010, p. 69). Although nothing is every truly "permanent," it is not difficult to see that a digital image of a book is in many ways less fragile than its print original.

The second benefit of digitization is space. A digitized collection means that a library can remove duplicate printed items, reduce the total amount of printed items, and eliminate many of the costs associated with maintaining large print collections. One library that demonstrates the radical concept of space conservation is in Texas; this library has fully embraced e-book technology. In September 2013,

Bexar County launched BiblioTech, the first all-digital public library in the United States (Cole, 2014). Bexar County, with a total population of 1.7 million, has been experiencing explosive growth outside the city limits of San Antonio, with the result that residents are farther away from the city's branch libraries (Cole, 2014). Within the library's mere 4,800 square feet, it offers classes, school tutoring, and children's reading programs, and provides meeting space, study rooms, a reading lounge, a community room, and also a café (Cole, 2014). How is the library able to offer so much in such a small space? The answer is by eliminating the print books. Instead, the library circulates 600 basic e-readers and 200 enhanced e-readers pre-loaded with children's content; inside the library are 48 desktop computers, 40 iPads, and 10 laptop computers (Cole, 2014). Best of all: patrons have access to over 25,000 e-books (Cole, 2014). What is most telling is that this library still has a physical space, albeit a smaller one than most libraries. In creating the physical space, the planners recognized that a library is far more than a repository for books and serves a vital social function (Cole, 2014).

Yet another benefit of digitization is the hope that it will increase access for the digitized works. This encompasses the obvious: users no longer need to physically visit a library to view items, and multiple users can view a single digital copy of an item. Digitization also increases the usability of materials that are otherwise inaccessible due to their fragile condition (Digital Public Library of America, 2015). This concept also encompasses something more subtle: digitization is ending the distinction between circulating libraries, meant for public readers, and research libraries, meant for scholars and often closed to the public (Cohen, 2009). The reality is that not just anyone from the public can waltz into Harvard University's library (Cohen, 2009). The concept of access also encompasses the ability to obtain search results on Google Books for all types of books, be they in print, out of print, public domain, or in copyright. According to the Online Computer Library Center, less than 20 percent of books are in print and only 20 percent are in the public domain (Schmidt, 2005). According to Google, this leaves 60 of the world's printed books hidden from the public (Schmidt, 2005).

From an environmental perspective alone, e-books have many advantages over traditional paper books. A single e-reader can replace many books, e-book online distribution has less of an environmental impact than the distribution system for paper books, users can greatly reduce their carbon footprint by reading e-books instead of paper, and e-books have the ability to reach the market much more quickly than print books (Manley and Holley, 2012, p. 306). Unlike physical books, digital copies do not wear out, which eliminates the need for libraries to reorder heavily used popular titles (Woo and Trachtenberg, 2011).

There are many social benefits of e-books as well. People already spend more time in front of screens and less time in front of printed books, and thus e-books may promote reading; e-books are faster and

cheaper to produce than paper books and are often cheaper to buy; e-books are easily updateable; e-books are searchable; e-books are portable, and thus a reader can carry an entire library on a single e-reader; e-books can be delivered almost instantly; e-books can be annotated without harming the original work; e-books make reading accessible to people with disabilities; and e-books can be hyperlinked for easier access to additional information (Simonson, 2015, p. 49). The argument that e-readers are expensive and thus available to only those who have the means to purchase them is weakened by the fact that many libraries now offer e-readers on loan.

III. CONTRACT ANALYSIS

A. Analysis Methods

For the purposes of this paper, four contracts were located and reviewed. Initially, Google was extremely secretive regarding its contracts with the library partners (Leetaru, 2008). Public dissent soon became apparent, and several arguments emerged that because the universities are state-funded, their legal agreements should be public record (Leetaru, 2008). This led to some of the library partners eventually publishing their contracts online (Leetaru, 2008). After determining which libraries and universities were Google Books Library Partners, the author visited the Google Books blog and the libraries' and universities' websites to determine whether any of the contracts were available online; the author located four contracts. The fifth contract discussed in this paper, Stanford University's, was not located online. However, the Second Circuit discussed the contract in its opinion at length; this discussion was sufficient to provide insights for the purposes of this paper.

As previously mentioned, there is scant scholarly literature examining the contracts between Google and the Library Partners. Thus, after performing the literature review, the author chose to focus on the contract analysis by sourcing the contracts, reviewing their provisions, and comparing the provisions of each contract to the others. The contract analysis benefitted from the author's legal background. The contracts appear to be based on a standard template, leading to the conclusion that the template was likely created by Google. It appears that the parties modified the template to suit their needs. The author reviewed each contract and paid particular attention to the following provisions: (1) the contracting parties (Google Inc. for the American contracts and Google Ireland Limited for the British contract, for example), (2) choice of law provisions (The British Library contract specifically states that it is governed by English law), (3) the various time limits within each contract, (4) how the costs are allocated among the parties, and (5) the breadth of each party's right to use the digital copies, whether for profit or otherwise. In the following sections, particular attention was also paid to the "fairness" of the contracts. The concept of "fairness" is susceptible to being viewed as subjective, yet two principles hopefully lend

some objectiveness: a contract may be considered unfair by the way in which it was created, such as by undue influence; or a contract may be considered unfair because one or more terms is more favorable to one party than the other (Thomas, 2010, p. 177).

B. Specific Contract Provisions

Google asserts that pursuant to its contracts with the libraries, each library receives a digital copy of every scanned book to preserve and, where copyright law allows, make available to patrons (Google Inc., 2017a). In reality, the terms are much more complicated. As a broad generalization and as articulated by the Second Circuit, Google allows each library to download copies of both the digital and machine-readable versions of the books that the library submitted to Google for scanning (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). This only applies to the books that the particular library itself provided to Google; the contracts do not give permission for a library to download digital copies of books that another library submitted (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). Google gives all participating libraries access to the Google Return Interface ("GRIN") in order to facilitate the libraries' downloading of their digital copies (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210).

The agreements between Google and the libraries require the libraries to abide by copyright law in utilizing the digital copies they download and to take precautions to prevent dissemination of their digital copies to the public (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). The contracts do not specify what precautions the libraries should take. For example, the University of California's and the Committee on Institutional Cooperation's contracts require the universities to implement technological measures to restrict automated access to the digital copies, and reasonable steps to prevent third parties from downloading and distributing the copies or any parts of them (Google Inc., 2007b, p. 9), but there are no specifics about how to do so. In the lawsuit, the plaintiffs raised concerns that Google's distribution of the digital copies to the libraries exposes the plaintiffs to risks of loss if the libraries use their digital copies in an infringing manner, or if the libraries fail to maintain security over their digital copies with the consequence that the digital books may become freely available as a result of hackers (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 228). However, the Second Circuit rejected these arguments on two bases (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 228-229). First, the court believed that the plaintiffs are adequately protected via existing law. Although the court acknowledged the possibility that libraries could use the digital copies Google created for them in an infringing manner, the libraries themselves could be held liable to the plaintiffs for such infringement (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 229). The court also acknowledged the possibility that in such a lawsuit, the plaintiffs could present evidence that Google was aware of or encouraged the libraries'

infringing practices, in which case Google could be liable as a contributory infringer (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 229). Second, the court believed it a remote possibility that the libraries would actually engage in copyright infringement or allow the digital copies to be hacked. Despite the acknowledgement of the possibility of infringement, the court found it to be sheer speculation that the libraries would misuse their digital copies or that Google would be liable as a contributory infringer (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 229). The court further found that although there is the additional possibility that the libraries might incur liability by the negligent mishandling of, and failure to protect, their digital copies, leaving them unreasonably vulnerable to hacking, that, too, is nothing more than sheer speculation (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 229). The court noted that the libraries are contractually obligated to protect against hackers (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 229). In sum, in the unlikely and purely speculative event of infringing behavior, the court believed that there were adequate legal safeguards in place to protect the plaintiffs' rights.

Although the contracts allow the libraries to utilize their digital copies in a non-infringing manner, it is interesting to note the rights that Google maintains in the digital copies, especially after termination of the contract. In general the contracts contain a post-termination clause which specifies that Google will continue to host and index the content after termination of the agreement for the purposes of providing search results (Laing, 2006, p. 137). Combining this with the fact that many of the contract provisions survive termination, the ultimate effect may be that the rights that exist in the newly created digitized files belong to Google, not the library that provided the books to Google to be digitized (Laing, 2006, p. 137). It should be noted that the actual copyright holder's contractual rights are unaffected, not least because the copyright holder was likely not a party to the contract with Google (Laing, 2006, p. 137). However, as time goes on and books fall out of copyright, the value of Google's rights in the digitized files will only increase (Laing, 2006, p. 137).

C. Specific Contracts

i. University of California and Committee on Institutional Cooperation

In 2006, the University of California signed a six-year contract with Google which anticipated the digitization of at least 2.5 million books (Albanese, 2006, p. 14). The following year, the Committee on Institutional Cooperation (now called the Big Ten Academic Alliance) also signed a six-year contract with Google. The Big Ten Academic Alliance is a consortium of 12 research universities with a total of 13 libraries; the members are the University of Michigan, the University of Wisconsin, the University of Minnesota, the University of Chicago, the University of Illinois, Indiana University, the University of

Iowa, Michigan State University, Northwestern University, Ohio State University, Pennsylvania State University, and Purdue University (Google Inc., 2007b, p. 1). The contract calls for the digitization of not less than 10 million of the universities' 79 million volumes (Google Inc., 2007b, p. 3). Except as noted below, the terms of both contracts are substantially similar.

Each university is responsible for the costs of (1) locating, pulling, moving, and re-shelving the books, (2) university employee labor, (3) network bandwidth and data storage required to receive the digital copies, (4) any conservation efforts needed prior to digitization, and (e) barcoding and associated data entry (Google Inc., 2007b, p. 4). Google is responsible for the costs of (1) Google employee labor, (2) the scanning hardware and software, (3) the scanning facilities, (4) transportation of the books to and from the scanning facilities, and (5) insurance to cover the books (Google Inc., 2007b, p. 4).

The contracts grant Google broad access to use the digitized copies as part of Google's services, including all aspects of Google Books (such as indexing the works, displaying the works, and making snippets available) (Google Inc., 2007b, p. 6). For public domain works or copyrighted works where Google has obtained permission, Google can provide, license, or sell its digital copies (Google Inc., 2007b, p. 5). Although the Second Circuit downplayed Google's commercial motive with regard to the Google Books project, one cannot ignore the commercial benefits Google retained for itself within the contracts.

Google's rights to use the digital scans are in direct contrast to the universities' rights to do so. As part of the contracts, each university receives one copy of each scan (Albanese, 2006, p. 14). This is perhaps the most controversial aspect of the contracts. Each university can use the digital copies as part of services offered on the university's website to advance the university's "academic purposes to support the instruction, research, and scholarship of its campus community," subject to copyright laws (Google Inc., 2007b, p. 9). Such use must be "academic"; notably, there is no such restriction on Google's rights to use the scans. The universities cannot receive any payment for the digital copies except that they can recover from end users the costs of printing and copying (Google Inc., 2007b, p. 9). Again, this is in direct contrast to Google's rights to provide, license, or sell its digital copies. Since Google uses the digital scans for text mining, for its search engine, and for its translation software, it is perhaps not surprising that Google included these prohibitions against the libraries' commercial use in the contracts (Ruiz, 2011). Yet such perceived inequalities have not gone unnoticed. Several critics have pointed out that unlike Google, the universities cannot share, license, or sell their scans to any third party (Albanese, 2006, p. 14). In effect this results in Google's being the only entity to have the most comprehensive collection of the digitized works, leading some to conclude that it is "a private library system controlled by a single corporation" (Carlson, 2006, p. 32).

The University of California's distribution rights are further defined by the copyright status of the original work. The university is free to distribute works not covered under copyright to scholars and students for research purposes (Carlson, 2006, p. 32) – notably, limited to "research" purposes. However, there are limits on copyrighted material (Carlson, 2006, p. 32). The university can redistribute no more than 10 percent of scanned material to other libraries or schools, even for educational purposes, which limits its involvement in interlibrary loans (Albanese, 2006, p. 14). Before receiving the digital copies, other institutions must enter a separate written agreement with Google regarding the use of the copies and provide indemnity to Google in the event of litigation (Carlson, 2006, p. 32).

The Big Ten Academic Alliance contract is the first time that Google contracted not to provide the participating libraries with copies of their digitized copyrighted books (Albanese, 2007, p. 18). Instead, Google will hold such copies in a complex escrow arrangement. Google will only release digital copies held in escrow to the universities under the following circumstances: (1) the work becomes in the public domain, (2) the copyright holder has granted permission, (3) "well established case law exists" that copyrighted works can be copied and held by the universities without infringing on the rights of a copyright holder, (4) Google is in material breach of its obligations and does not remedy such breach within 90 days, or (5) Google and the universities agree in writing that the release of the copyrighted work is legally supported and appropriate (Google Inc., 2007b, p. 8).

Finally, never one to miss an opportunity to promote its brand, the contracts also require the universities to identify all publicly available digital works as "Digitized by Google" (Google Inc., 2007b, p. 9). This is arguably yet another example of Google's commercial motivation that the Second Circuit downplayed.

ii. The University of Texas at Austin

The University of Texas at Austin and Google entered into a six-year contract in January 2007 which was amended in June 2009 (Google Inc., 2007a, p. 1). The contract is governed by Texas law (Google Inc., 2007a, p. 12). The terms of the contract are substantially similar to those of the University of California contract discussed supra.

The University of Texas at Austin was one of the institutions that limited its digitization with Google to only specific collections (Howard, 2012). The university agreed that Google could digitize its Latin American collection, which consists of approximately 500,000 volumes (Howard, 2012). According to Fred Heath, vice provost and director of the University of Texas Libraries, the university was not interested in choosing from its 10 million volumes in all of its libraries, shipping those materials to Google, and then re-shelving them once they returned to the university (Howard, 2012). Further,

choosing a specific collection allowed the work to be done much more quickly than the university could have done by itself. The university estimated that it would take a hundred years to digitize the Latin American collection on its own; Google apparently completed the work in two years with no loss of materials (Howard, 2012).

iii. The British Library

The British Library partnered with Google Ireland Limited in 2011 (Ruiz, 2011). The contract is for a period of six years (Google Ireland Limited, 2011, p. 8). Unlike its American counterparts, this contract is governed by English law (Google Ireland Limited, 2011, p. 11).

Interestingly, the British Library's contract with Google does not specifically set forth that it is non-exclusive. However, this is surely part of the contract as the British Library's own website divulges its scanning activities with other partners: "Recently, the Library announced a partnership with brightsolid to digitise up to 40 million pages of its newspaper collections and previously the Library partnered with Microsoft to digitise 65,000 19th century books, some of which are now available as an App on Apple's iPad" (British Library, 2011).

The British Library's contract requires it to provide no less than 250,000 volumes to Google for digitization (Google Ireland Limited, 2011, p. 3). The division of expenses is similar to the University of California contract: the library will pay for locating, pulling, and re-shelving the materials, network bandwidth and data storage to receive its digital copies, any conservation required prior to digitization, and creation of standard metadata, and Google will pay for transportation to the digitization facilities, digitization, and insurance (Google Ireland Limited, 2011, p. 4).

Similar to the American contracts reviewed above, section 4.3 grants Google broad access to use the digitized copies as part of Google's services, including all aspects of Google Books (such as indexing the works, displaying the works, and making snippets and bibliographic information available) (Google Ireland Limited, 2011, p. 5).

In contrast, sections 4.7 and 4.8 of the contract set forth the British Library's rights with respect to the digital copies. Regardless of the works' copyright status, the library can use the copies as part of noncommercial services offered on its website (Google Ireland Limited, 2011, p. 6). The University of California's and the Big Ten Academic Alliance's contracts use the word "academic" and the British Library's contract uses the word "noncommercial" but the result is the same – Google can profit from its use of the digital scans but the libraries cannot. This is further emphasized by the provision prohibiting

the British Library from sharing, licensing, or selling the copies to any third party (Google Ireland Limited, 2011, p. 6).

Specifically with regard to public domain works, the library may provide the digital copies to other institutions solely for noncommercial research, scholarly, or academic purposes (Google Ireland Limited, 2011, p. 6). This is similar to the University of California contract, which allows it to distribute public domain works to scholars and students for research purposes. Prior to doing so, however, the receiving institution must enter into a written agreement with Google, to ensure that the institution will not distribute the digital copy or utilize it in a way similar to Google Books services (Google Ireland Limited, 2011, p. 6). This is even more restrictive than the University of California contract, which required such a written agreement for copyrighted works only.

As with the other contracts reviewed, the British Library must implement measures to restrict automated access to the copies, and use best endeavors to prevent third parties from downloading and distributing the copies or any parts of them (Google Ireland Limited, 2011, p. 6).

Google may elect not to digitize all or some of the selected materials whether due to cost issues, conservation concerns, or otherwise, and Google absolves itself of liability for providing the digital copies to the British Library as a result (Google Ireland Limited, 2011, p. 10). Google also reserves the right to suspend performance or terminate the contract if it reasonably determines that it is commercially impractical to continue performing its obligations (Google Ireland Limited, 2011, p. 8). These provisions are notable in that they are so one-sided. By carving out these rights for itself, it appears that Google could cease its scanning activities midway through the project, leaving the British Library to pay for the costs of collecting and re-shelving the books that were in Google's possession. Further, in such a scenario, the British Library would be forced to seek out another digitization contractor to complete the work – again, at its own expense.

iv. Stanford University

Stanford was one of the original five libraries to partner with Google in December 2004. Although a copy of its contract was not located for this paper, the Second Circuit discusses the contract in its opinion. The Second Circuit opined that Google's agreement with Stanford University appears to be less restrictive on the university than its agreements with other libraries (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 230). The contract ostensibly permits Stanford's libraries to "provide access to or copies from the Stanford Digital Copy" to a wide range of users, including individuals authorized to access the Stanford University Network, individuals affiliated with "partner research libraries," and "education,

research, government institutions and libraries not affiliated with Stanford," and to permit authorized individuals to download or print up to 10 percent of Stanford Digital Copy (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 230). On the other hand, the agreement requires Stanford to employ its digital copies in conformity with copyright law (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 230). Without evidence to the contrary, the court found it reasonable to construe these potentially conflicting provisions as meaning that Stanford may do the enumerated things ostensibly permitted only to the extent that doing so would be in conformity with copyright law (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 230).

IV. DISCUSSION

A. Contract Best Practices

It goes without saying that any entity must carefully weigh the pros and cons of entering into a contract with another entity, and ensure that it has the optimal tools for negotiation. This is particularly true for a library that is contemplating partnering with Google as part of the Library Project, for not only are there serious concerns about the quality of the project, it may be that Google has far greater negotiating ability than a small library. A review of some of the contracts between Google and the Library Partners reveals important things that libraries should consider. Such considerations apply both to libraries seeking to contract with Google for digitization as well as libraries wishing to contract with another third-party, for-profit company. Although some of these issues could apply to any library contract, the following list is meant to guide libraries specifically with regard to digitization.

i. Deciding to Digitize

The first question, of course, that any library must consider is deciding whether to digitize or not. The Online Computer Library Center recently released a report concerning the state of digitization among American libraries; 92 percent of the public libraries surveyed have unique and locally significant materials (Morgan and Proffitt, 2017, p. 10). That same report, however, found that only 37.6 percent of libraries had engaged in digitization activities in the past three years (Morgan and Proffitt, 2017, p. 6). Some institutions, however, make digitization a priority: the Library of Congress, through its new leader, Carla Hayden, "has committed to opening up public access to its collections through digitization" (Rosenberg, 2017).

ii. Disclosure

Libraries should anticipate that their digitization contracts may become public. For example, the University of California's contract began as confidential, but was soon released in part as a response to an open-records request from The Chronicle newspaper (Carlson, 2006, p. 32). The British Library's contract was quickly made public after a Freedom of Information Act request (Ruiz, 2011). On the other hand, other institutions embrace transparency as one of the principles they are guided by during the contract negotiation process. The Library of Congress assumes that all of its third-party digitization agreements will be nonconfidential, public information (Library of Congress, n.d.). In 2007, the United States National Archives publicly revealed its general plan for digitization as well as for a prospective business deal, with the understanding that public comment would involve the community, potentially improve the results, avoid any public relations problems, and earn admiration for its openness (Kaufman and Ubois, 2007).

iii. Non-Exclusivity

Google asserts that its contracts with all libraries are non-exclusive, which means that libraries can continue their own scanning projects or work with others while they work with Google to digitize their books (Google Inc., 2017a). Obviously this is in the libraries' best interest as it allows them greater flexibility with regard to their digitization projects. Libraries should ensure that their contracts allow them to engage other third-party vendors for digitization as well as taking on digitization activities in-house.

iv. Time Limits

Time limits are an important part of any contract. Most of Google's contracts with the Library Partners specify that the digitization activities are for a period of six years with the option to renew for additional years. Before a library agrees on a timeframe, it should consider the length of time that will be required to deliver a quality product. As previously mentioned, one of the criticisms of the Google Books project has been the haste in which the digitization was done, likely contributing to the many errors in the scanning and metadata.

The University of California contract specifies that any books pulled for scanning are to be back on the libraries' shelves within 15 days (Carlson, 2006, p. 32). The other reviewed contracts are silent as to any timeframe for return of the books, perhaps because Google was finding it difficult to adhere to such time constraints.

Unfortunately for the libraries, the American contracts reviewed provide no time limit on the libraries' restrictions for using the digital copies. Libraries should endeavor to negotiate away from these restrictions. "Exclusivity on the right to manage and distribute the digital content should be limited to the term of the contract" (Kaufman and Ubois, 2007). Yet some contracts, such as British Library's, do specify such a time period. All restrictions on the British Library's use of its digital copies expire after 15 years (Google Ireland Limited, 2011, p. 6). There are some, however, that believe this is too long. Around the time of the British Library's contract negotiation with Google, the European Union issued a report about the digitization of culture and set a maximum of seven years for the most optimal arrangements involving partnerships between public and private entities (Ruiz, 2011). The European Union meant this time limit to strike a balance between the interests of businesses and public institutions (Ruiz, 2011). Further, the British Library's contract with Google was unique in that most digitization contracts between the British Library and the National Archives last for ten years (Ruiz, 2011). Thus, there is literature criticizing the Google contract's 15 year timespan and instead recommending a cap of seven years (Ruiz, 2011). The Library of Congress recommends a time period of less than three years for any restrictions upon the Library's right to distribute the digital copies externally, noting that any embargo period is generally disfavored (Library of Congress, n.d.).

v. Costs

One of the most important considerations is cost. In fact, the Library of Congress, for example, specifically limits its third-party digitization projects to those governed by no-cost contracts (Library of Congress, n.d.). Google's contracts with the Library Partners generally split the costs in half: the library pays for locating, pulling, and re-shelving the materials, network bandwidth and data storage to receive its digital copies, any conservation required prior to digitization, and creation of standard metadata, and Google pays for transportation to the digitization facilities, digitization, and insurance. Some organizations, like the Center for Research Libraries, require third-party vendors to be responsible for any preservation costs prior to digitization (Center for Research Libraries, 2005).

In addition to the costs of digitization itself, libraries should also consider the costs to store the electronic copies. Although the costs of storage space have decreased in recent years, "digital projects are storage hungry by their very nature" (Hall, 2011), and thus costs can quickly add up.

vi. Breadth of Digitization

Libraries should consider what materials they actually want included in the digitization project. The University of Texas at Austin and the British Library, for example, limited their contracts with Google to

only specific collections. Since the Google contracts require the libraries to pay for the costs of locating, pulling, and re-shelving the materials, libraries may opt to limit their digitization to manageable collections or portions of collections.

Even though the British Library has been involved in digitization efforts since the early 1990s and has "a three-figure list" of ongoing digitization projects, it estimates that only less than one percent of its materials have been digitized to date (Hall, 2011). The contract with Google for digitization of 250,000 books may sound like a lot, but even the British Library says that this amount is "relatively modest" compared to the 15 million books in its possession (Hall, 2011). There is no doubt that Google currently possesses the most ability for large-scale digitization, but the British Library acknowledges that it has many smaller initiatives, often built around a single item in the library or having a sponsor or benefactor interested in making certain documents available to a larger audience (Hall, 2011).

On the other hand, digitizing large collections has its advantages as well. For example, when the British Library contracted with Microsoft in 2007, it elected to digitize its entire 19th century literature collection, unlike previous digitization projects where material had been selected on an item-by-item basis (Ashling, 2007). Being nonselective has various advantages: it widens access, it lessens the domination of well-known authors, the works of relatively unknown writers "will be brought to the attention of scholars as easily as material by Charles Dickens," and entire shelf runs can be removed for scanning at one time (Ashling, 2007).

vii. Copyright Issues

At the heart of the Google Books lawsuit, of course, was the issue of digitization of copyrighted works. Even though Google's actions were found to be legal in the United States, some libraries may wish to restrict their digitization efforts to only works in the public domain. This is particularly true for libraries outside of the United States, to which the Second Circuit's "fair use" holding does not reach. The British Library, for example, is "keen" to digitize 20th century material, but is currently unable to because of copyright restrictions (Hall, 2011).

Even though the Google Books case allows American libraries to digitize copyrighted works, some libraries choose not to. For instance, the Library of Congress refuses to engage with third-party digitizers for anything but public domain works. The Library of Congress (n.d.) has delineated its own requirements for digitized materials; they must be (1) in the public domain, (2) materials for which, in the Library of Congress' judgment, there are no known copyright restrictions, or (3) materials for which the

digitizing partner agrees to obtain copyright permissions for both its activities and the Library of Congress' activities with respect to the project.

Many of Google's partners, throughout the course of the Google Books litigation, elected to limit their scanning to only works in the public domain; such institutions include the Complutense University of Madrid, the National Library of Barcelona, Princeton Library, the Bavarian State Library, Keio University, Columbia University Libraries, the Municipal Library of Lyon, the National Libraries of Florence and Rome, the Austrian National Library, the National Library of the Netherlands, and the British Library.

viii. Metadata

Libraries must decide whether they expect to receive metadata about the digitized items from the digitizer itself. The Library of Congress, for instance, requires third-party digitizers to provide associated core metadata sufficient to make the digitized copies usable by the Library of Congress, and if the digitizer creates enhanced metadata for its value-added features, the Library of Congress encourages, but does not require, the digitizer to provide that metadata as well (Library of Congress, n.d.). Interestingly, the contracts that Google entered into for the Google Books project are just the opposite: they require the library to provide metadata about the item. In the wake of Google's blundering of metadata (i.e. rampant errors in the millions), the libraries that have already partnered with Google may well be breathing a sigh of relief that Google's provision of metadata was not part of the deal.

ix. Data Mining Capabilities

At least as – or arguably even more – valuable as the digital scans themselves is the ability to perform word searches within the scans. Undoubtedly Google sees this as the most enticing aspect of Google Books, since word searches form the backbone of the entire project. Libraries should determine whether the third-party digitizer will provide only digital scans of the documents or whether it will also convert the documents into digitally-readable text. As explained by Richard Boulderstone, Director of E-Strategy and Information Systems at the British Library: "The scanning process is similar to creating a picture book, which is just taking images. But as a picture book it's not that useful, as that does not enable key word searches" (Hall, 2011). Some of the British Library's digitization projects have not contributed very much to the digital collection, but the British Library believes their value is in their machine-readable text, providing an important tool for researchers seeking to data mine the digital collection (Hall, 2011). For some projects, the British Library works with a third-party OCR specialist (Hall, 2011). The process of converting a document into a digitally-readable text is more than just character recognition; it involves

image pre-processing, image cleaning, and analyzing the structure of images (Hall, 2011). Sometimes characters must be read individually and then the entire document must be re-assembled using a dictionary (Hall, 2011).

x. The Number of Digital Copies the Library Receives

The contracts that Google entered into with its Library Partners are very specific in allowing the libraries only a single digitized copy of each scanned item. For some libraries this may be sufficient. Other libraries may wish to seek additional copies during the contract negotiation process. As part of the Library of Congress' guidelines for digitization contracts, one mandate is that the Library receive one digital copy suitable for archival purposes; the Library "generally expects to receive an additional copy, in a format to be agreed upon, for purposes of providing access to Library patrons" (Library of Congress, n.d.). Libraries may wish to "request a digital copy to preserve themselves, further safeguarding materials by preserving them in multiple locations" (Association for Library Collections and Technical Services, 2013).

xi. The Library's Use of the Digital Copies on its Website

One purpose of digitization is to open up libraries' collections to both their patrons and the general public, so one goal of digitization contracts should be the libraries' ability to display the digital copies on their websites. Google's contracts with the Library Partners are commendable in that they allow the libraries to display the digital copies on their websites immediately, for academic or noncommercial purposes. Lest this seem like a minor issue, it should be noted that even the all-powerful Library of Congress is not always able to obtain this in its digitization contracts, and lists it as a "preference" within its digitization guidelines: "The Library prefers to have the immediate right (but not obligation) to provide full and unrestricted access to the digitized copies on its Web site" (Library of Congress, n.d.).

xii. The Library's Use of the Digital Copies for Inter-Library Loans

Libraries should review any potential contracts for limitations that may be imposed on inter-library loans of the digital scans. The University of California, for example, can redistribute no more than 10 percent of scanned material to other libraries or schools, even for educational purposes, which limits its involvement in interlibrary loans (Albanese, 2006, p. 14). Before receiving the digital copies, other institutions must enter a separate written agreement with Google regarding the use of the copies and provide indemnity to Google in the event of litigation (Carlson, 2006, p. 32). This is a potentially huge obstacle for libraries, which routinely participate in inter-library loan programs and are not accustomed to

being required to enter into written agreements with third-party, for-profit corporations before receiving such inter-library loans.

xiii. The Library's Commercial Use of the Digital Copies

Unlike Google, the Partner Libraries cannot share, license, or sell their scans to any third party. Any library contemplating a digitization contract with a third party must carefully analyze whether this is an acceptable term.

Interestingly, the British Library cites revenue generation as one of its goals of digitization in general ("By digitising our collection we will . . . Generate income to help sustain our long-term digitisation programme") (British Library Board, n.d.). Such commercial profit would violate the terms of the British Library's contract with Google, so perhaps such action is allowable under its other contracts or if and when the British Library performs in-house digitization. The British Library advises that its digitized materials may be "owned by a commercial partner of the Library" (British Library Board, n.d.), which could also be a source of potential income generation.

Although the New York Public Library is not representative of most libraries in light of its sheer size and popularity, its embrace of revenue generation with regard to digitized items is worth noting. The New York Public Library recently changed its protocol regarding downloading digital images. As of April 17, 2017, patrons can download select files that are marked as having no known United States copyright in the highest resolution the Library has to offer, for no fee (New York Public Library, 2017). Such images can be used, reused, and shared commercially and otherwise, yet may still be subject to rights of privacy, rights of publicity, and other restrictions (New York Public Library, 2017). The New York Public Library profits from selling downloads of images still under copyright as well as selling t-shirts, note cards, mugs, tote bags, and other items containing hand-picked images from its collections through the third-party vendor Zazzle (New York Public Library, 2017). The New York Public Library is continuing to add new images for sale on a regular basis (New York Public Library, 2017).

xiv. The Digitizer's Use of the Digital Copies

A cornerstone of every contract Google enters into with a Library Partner is Google's right to use the digitized copies on the Google Books website. In fact, this is arguably the most valuable part of the contract from Google's perspective, furthering its mission "to organize the world's information" (Google Inc. 2004). This right is succinctly delineated in Google's contract with the British Library, for example, which grants Google broad access to use the digitized copies as part of Google's services, including all aspects of Google Books (such as indexing the works, displaying the works, and making snippets and

bibliographic information available) (Google Ireland Limited, 2011, p. 5). Although it is easy to jump to the conclusion that allowing Google broad access to use the digital copies on its website is a bad thing and is further proof of Google's motivation for profit, perhaps there is some benefit. The Library of Congress, in fact, prefers that its third-party digitization partners "provide some form of free public access to the digital copies," and the Library "will not restrict the partner from charging for copies of and/or access to the materials, particularly when the partner has developed value-added features for search, discovery, and display" (Library of Congress, n.d.).

xv. The Library's Remedies in the Event of Discontinuation of the Project

In the British Library contract, there is an argument that Google could cease its scanning activities midway through the project, leaving the British Library to pay for the costs of collecting and re-shelving the books that were in Google's possession. Further, in such a scenario, the British Library would be forced to seek out another digitization contractor to complete the work at its own expense. Libraries should thus ponder whether this is an acceptable term in their own contracts. If not, a provision could be added to the contract which delineates remedies in the event of the digitizer's breach.

xvi. User Privacy

As highlighted earlier, the issue of privacy is especially relevant for libraries, because Google and other third-party organizations may have vastly different standards for handling user information than libraries do. Internet services often make broad use of patron, usage, and circulation data "as a matter of course" (Kaufman and Ubois, 2007). Despite the importance of this issue and the fact that numerous laws govern the collection, use, and protection of patron information and other personally identifiable information (Kaufman and Ubois, 2007), the contracts reviewed between Google and the Library Partners are silent with regard to this issue. Libraries should be prepared to question the third-party vendor about "the collection, ownership, sharing, protection, and use of log file data" (Kaufman and Ubois, 2007). Libraries should also question the vendor to ensure that personally identifiable information (along with information that could be used to deduce personally identifiable information, such as IP addresses) is secured and removed from log files (Kaufman and Ubois, 2007).

xvii. The Library's Security Measures

As previously mentioned, Google's contracts with the Library Partners require the libraries to abide by copyright law in utilizing the digital copies they download and to take precautions to prevent dissemination of their digital copies to the public (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 210). This highlights the need for libraries to create and maintain effective and robust internet use

policies. For example, libraries in the United Kingdom have implemented various measures to minimize copyright infringement, such as authentication, monitoring use, crafting an appropriate internet use policy, installing filtering software, preventing patrons from downloading software that could be used to infringe, and displaying copyright notices (Spacey et al., 2014).

If a library permits a work to be read online, it should take efforts to ensure that inappropriate levels of access do not take place (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100). Libraries could consider using a mass download prevention system called "choke" to measure each user's rate of activity (such as the rate of pages read) (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100). In such a case, if a user's rate of activity exceeds certain thresholds, the system assumes that the user is mechanized (e.g., a web robot) and blocks that user's access for a set period of time (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100).

As part of the above discussion regarding potential copyright infringement, libraries must also review whether they have implemented adequate security measures to ensure against hacking. Of course, few libraries have the means to implement Google-like security measures. However, this does not mean that libraries should ignore or downplay internet security, especially in light of the Second Circuit's attention to this issue.

xviii. The Digitizer's Security Measures

The Second Circuit highlighted the many ways in which Google protects the digital copies under its control: the digital scans are stored on computers walled off from public internet access and protected "by the same impressive security measures used by Google to guard its own confidential information" (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 228). In fact, during the course of the litigation the plaintiffs' own security expert praised these security systems, remarking that "Google is fortunate to have ample resources and top-notch technical talents" that enable it to protect its data (*The Authors Guild, Inc. v. Google Inc.*, 804 F.3d 202, 228).

In a case related to *The Authors Guild* case, the Second Circuit acknowledged HathiTrust's more than adequate security measures (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100). Such measures include physical security controls: servers, storage, and networking equipment are mounted in locked racks, and only six individuals have keys (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100). The data centers housing the servers, storage, and networking equipment at each site are monitored by video surveillance, and entry requires use of both a keycard and a biometric sensor (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100). Such measures also include system security controls: network access is

highly restricted (even for the staff), and two levels of network firewalls are in place at each site (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100). Finally, such measures include internet security controls: access to the service is governed by the HathiTrust rights database, which classifies each work by presumed copyright status, and also by a user's authentication to the system (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 100).

B. Alternatives for Libraries

There is every indication that Google is slowing down the Google Books project, at least as to its scanning and promotion activities, which had their peak around 2010 and 2011 (Rosenberg, 2017). One journalist recently declared that "Google Books has settled into a quiet middle age" and has "lost its drive and ambition" (Rosenberg, 2017). Some speculation is that the settlement agreement Google crafted with the plaintiffs would have opened up the Library Project to extensive commercialization and a potential for revenue generation (Murrell, 2016). When the court rejected that agreement in March 2011, it cut off that potential source of revenue (Murrell, 2016). Since then, Google's scanning activity has drastically reduced, the Google Books blog was discontinued in 2012, the Google Books Twitter feed went silent in 2013, and the Google Books staff has mostly left or was reassigned (Murrell, 2016). Google's work on the Google Books project is now more behind the scenes, such as perfecting the technology, improving the scanning machines, acquiring new content, processing the content properly so that entire books can be viewed online, and adjusting the search algorithm so that searches are more precise (Rosenberg, 2017).

The following examples are but a few that exist for libraries today who wish to have an alternative to Google for scanning, storing, and offering their digital collections. Although Google dominates in terms of sheer volume, there are advantages to the smaller, non-profit alternatives listed below. These entities are not subject to the "changing priorities of a gigantic technology corporation," and they have a consistent dedication to books, "unencumbered by distractions like running one of the largest advertising businesses in the world or managing a smartphone ecosystem" (Rosenberg, 2017).

i. In-House Digitization

In response to the Google lawsuit, some experts believe that scanning activities are cheap and easy and that books will be digitized without Google's involvement (Chu, 2009, p. A9). Whether digitization is actually cheap is subject to debate, yet there is no doubt that some libraries have begun to take control of scanning activities for themselves. For example, Emory University elected to go a different route rather than partner with Google. In 2007, on the same day that Google announced its partnership with the Big Ten Academic Alliance, Emory University announced a different approach to digitizing its collections

(Jaschik, 2007). The university purchased a Kirtas robotic book scanner, which can digitize as many as 50 books per day, turning each volume into a PDF file (Albanese, 2007). After scanning, the titles are uploaded to a website where scholars can access them and, if they wish, buy "print on demand" copies through Amazon, to help the university recoup some of its costs (Albanese, 2007). The university's director for digital programs, Martin Halbert, commented: "[T]his option allows us to retain much more control over digitizing our pre-1923 collection and reserving the right to freely make that digital content available to scholars and the public. The degree to which Google would have controlled the digitized versions of items in our collection was unacceptable, speaking as stewards of the intellectual assets of the library" (Albanese, 2007). Emory University's approach is seen as more library-driven (Jaschik, 2007). This alternative to Google is based on the idea that different libraries will want to digitize different materials, in significantly different scales, and that libraries will pick key holdings, not everything in their collections (Jaschik, 2007). Emory University, for instance, began its project by focusing on Southern history and culture rather than its entire collection (Jaschik, 2007).

ii. HathiTrust

HathiTrust is proof that even when libraries partner with Google for digitization, they have some options regarding collaboration and presentation. HathiTrust began in 2008 with the universities of the Committee on Institutional Cooperation (now the Big Ten Academic Alliance) and the University of California system to establish a repository to archive and share their digitized collections (HathiTrust, 2017), including materials digitized by Google. Since then HathiTrust has expanded to include books and journals from other sources such as the Internet Archive, Microsoft, and from libraries' in-house scanning projects (HathiTrust, 2017). However, its core remains the Google-digitized books (Murrell, 2016), and it has established mechanisms to accommodate the ingest of content from Google and the Internet Archive with no costs (HathiTrust, 2017). HathiTrust now has over 120 worldwide partners and contains over 15 million volumes (HathiTrust, 2017).

Like Google, HathiTrust and its member libraries were the subject of a lawsuit brought by The Authors Guild, Inc. and other plaintiffs for copyright infringement. One year prior to issuing its opinion in the Google Books case, the Second Circuit held that HathiTrust and the libraries' use of the digitized texts was "fair use" under United States copyright law (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 101). At issue was whether HathiTrust's use of copyrighted works constituted copyright infringement. As the court explained, HathiTrust permits three uses of the copyrighted works in its repository:

> Users can search for particular terms across all digital copies in the repository, and the search results show only the page numbers on which the search term is found within the text and the

number of times the term appears on each page (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91). Unlike the Google Books website, HathiTrust's platform does not display any text from the copyrighted work (either in "snippet" form or otherwise) (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91). Consequently, the user cannot view the page on which the term appears or any other portion of the book (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91).

Member libraries can provide patrons who have certified print disabilities access to the full text of copyrighted works (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91). A "print disability" is any disability that prevents a person from effectively reading printed material, including blindness or disabilities that prevent a person from physically holding a book or turning pages (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91). Such print-disabled users can obtain access to the digital works by using adaptive technologies such as software that converts the text into spoken words or that magnifies the text (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 91).

By preserving the copyrighted books in digital form, HathiTrust allows members to create a replacement copy of the work but only if the member already owned an original copy, the member's original copy is lost, destroyed, or stolen, and a replacement copy is unobtainable at a "fair" price elsewhere (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 92).

The court first held that the doctrine of fair use allows the libraries to digitize copyrighted works for the purpose of permitting full-text searches, because such activity is transformative (a word search is completely different than the book from which it is drawn), the libraries need to create a digital copy of the entire work in order to facilitate full-text searching, and full-text searching does not harm the market for the originals (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 97-99). The court also held that the doctrine of fair use allows the libraries to provide full digital copies of copyrighted works to their print-disabled patrons (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 103). Finally, the court declined to decide whether the doctrine of fair use allows the libraries to make a digital copy for preservation and remanded this issue to the district court to determine the plaintiffs' standing (*The Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 104).

iii. Internet Archive

The Internet Archive bills itself as a non-profit library of millions of free books, movies, software, music, websites, and more (Internet Archive, 2017). It is probably most well-known for its Wayback Machine, but it also houses an impressive number of digitized works as well as undertaking digitization activities

itself. The Internet Archive began its digitization project with a single scanning location in 2004, and has now expanded to include scanning operations at 33 locations in Asia, Europe, and North and South America (Hoffelder, 2013). To date, the Internet Archive offers over 11 million digitized books and texts (Internet Archive, 2017). This includes over 2.5 million public domain (pre-1923) books which are fully downloadable, and 500,000 modern (post-1923) books which are available to the blind and dyslexic and through the Internet Archive's lending system, Open Library (Kahle, 2017, p. 28). Open Library uses the same protection technologies (to ensure access to only one reader at a time) that publishers use for their in-print e-books distributed by such companies as OverDrive and Google Books (Kahle, 2017, p. 30).

iv. Digital Public Library of America

Ironically, Harvard University, which was one of Google's original library partners, is now one of the biggest proponents of an alternative to the Google Books project. The vision of a national digital library has been circulating among librarians, scholars, educators, and private industry representatives in the United States since the early 1990s (Digital Public Library of America, 2017b). Many institutions began digitizing their collections, but each collection stood alone (Digital Public Library of America, 2017a). In October 2010, 40 leaders from libraries, foundations, academia, and technology projects met at a national conference in Cambridge, Massachusetts, to create an alternative to Google Books called the Digital Public Library of America (Thompson, 2013, p. 53). The founders envisioned that this digital library would operate under the auspices of the Berkman Center for Internet & Society at Harvard University and would archive every book in the public domain and offer them online for free to anyone (Thompson, 2013, p. 53). Kenny Whitebloom, who initially managed the library project, explained that "this is a civic-minded engagement," and "there was a fear that Google Books was a commercial project, and this was an opportunity to create a digital library on its own terms" (Thompson, 2013, p. 53). Because the Digital Public Library of America does not have enough money to perform the digitization itself, instead it acts as an archive and has no holdings itself (Thompson, 2013, p. 53). The Digital Public Library of America operates as a search engine that coordinates with disparate American institutions that have created digital archives of their collections (Thompson, 2013, p. 53). The Digital Public Library of America took two years to create and officially launched in April 2013 (Digital Public Library of America, 2017b). The Digital Public Library of America ultimately wants to become a worldwide network that will bring together the holdings of all libraries and museums (Murrell, 2016). To this end, its technical infrastructure was designed to operate with Europeana, yet another response to the Google Books project (Murrell, 2016).

The Digital Public Library of America bills itself as more than just a search engine, explaining that its portal provides innovative ways to search and scan through the united collection of millions of items, including by timeline, map, format, subject, and partner (Digital Public Library of America, 2017a). Through some of its partners, called "service hubs," the Digital Public Library of America offers libraries a range of services: staff with expertise in managing digital collections, digitization equipment and software for reformatting, metadata creation or enhancement, file storage and backup, reasonable prices or help with locating funding, and community outreach programs to increase awareness of digital content of local significance (Digital Public Library of America, 2015).

Despite the strides the Digital Public Library of American has made, and the fact that public library involvement has doubled since 2014, only 6 percent of American public libraries are participating (Morgan and Proffitt, 2017, p. 7).

v. Europeana

Like the Digital Public Library of America, Europeana brings together the collections of various galleries, libraries, archives, and museums, but with a European rather than American focus (Europeana, 2017). The project began with the Bodleian Library, the University of Tartu Library in Estonia, and the Bavarian State Library (Kelley, 2011, p. 18(2)). The Bodleian Library contains Europe's largest collection of public domain digitized books and journals and more than a dozen special collections materials from Oxford (Kelley, 2011, p. 18(2)). The University of Tartu Library contributed autographs, photographs, maps, letters by figures such as Immanuel Kant, books, and video clips, and the Bavarian State Library contributed 3-D digital books and medieval manuscripts (Kelley, 2011, p. 18(2)). Europeana currently acts as a portal to more than 54 million items (Europeana, 2017). It appears that Google recognizes its power – within the British Library contract is a provision that nothing in the agreement is meant to restrict the library from allowing Europeana to crawl the digital copies' metadata (Google Ireland Limited, 2011, p. 7).

V. CONCLUSION

Although the Google Books project may be slowing down, the proliferation of contracts created during its heyday provides much insight for libraries. Analyzing these contracts with a critical eye has revealed terms that are beneficial to libraries and others that are not. This analysis has led to a framework that will hopefully assist libraries during the contract negotiation process. Once a library has decided to contract with Google or another third party for its digitization needs, the library must consider what contract terms best suit its needs. Some things libraries should consider are the confidentiality of the contract terms,

non-exclusivity provisions, time limits, the allocation of costs between the library and the digitizer, the breadth of digitization, whether the library chooses to digitize copyrighted works or whether it will limit digitization to public domain works, which party will be responsible for creating and providing metadata, whether the works will be available for data mining, how many digital copies the library will receive, the library's ability to use the digital copies on its website and for inter-library loans, the digitizer's ability to use the digital copies, both parties' rights to commercially profit from the digital copies, the library's remedies in the event of the digitizer's breach, user privacy, and the parties' responsibilities for security measures.

This research was successful in that it led to a comprehensive set of best practices for libraries. Yet this may prove to be just the tip of the iceberg. As Google slows down its digitization activities, other companies are ramping up, and libraries would benefit from future research that analyzes these other companies' contracts. Recommendations for future research include further analysis of the non-American contracts (as only one non-American contract, The British Library's, was located for this paper), analysis of libraries' contracts with other third-party companies and whether a set of best practices can be gleaned from them, and further analysis of libraries' role in Google Books in light of the fact that the project appears to be concluding.

BIBLIOGRAPHY

Albanese, A. (2006) Google scans, contract released; search firm also will link to archives, allow downloads. *Library Journal*, 131(16), pp. 14-15.

Albanese, A. (2007) Google adds 12 CIC libraries; Emory, eschewing the trend, decides to scan its own books. *Library Journal*, 132(12), p. 18.

Allsop, T. (2011) Beware Google geeks bearing gifts. *Information World Review*, 267, p. 11.

American Civil Liberties Union (2009) *Amended Google Book settlement: doesn't deal with privacy problems*, November 14. Available at: https://www.aclunc.org/blog/amended-google-book-settlement-doesnt-deal-privacy-problems (Accessed: February 14, 2017).

American Library Association (2005) *FAQ: USA Patriot Act*. Available at: http://www.ala.org/offices/oif/ifissues/issuesrelatedlinks/usapatriotactfaq (Accessed: February 20, 2017).

American Library Association (2017) *Authors for library e-books*. Available at: http://www.ala.org/transforminglibraries/a4le/learn-more (Accessed: January 29, 2017).

Arch, J. (2016) Ebooks in libraries: equal access to digital content? *Information Today*, 33(5), pp. 1, 24-25.

Ashling, J. (2007) Progress report: the British Library and Microsoft digitization partnership. *Information Today*, 24(10).

Association for Library Collections and Technical Services (2013) *Minimum digitization capture recommendations*. Available at: http://www.ala.org/alcts/resources/preserv/minimum-digitization-capture-recommendations (Accessed: April 16, 2017).

The Authors Guild, Inc. (n.d.) *Who we are*. Available at: https://www.authorsguild.org/who-we-are/ (Accessed: April 13, 2017).

The Authors Guild, Inc. (2011) Fourth amended class action complaint. *The Authors Guild, Inc. v. Google Inc.*, Case 1:05-cv-08136-DC, Doc. 985, October 14.

The Authors Guild, Inc. v. Google Inc., 804 F.3d 202 (2nd Cir., 2015), *cert. denied* 136 S. Ct. 1658 (2016).

The Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2nd Cir., 2014).

Bangeman, E. (2007) Google Book search adds libraries, books from University of Lausanne. *Ars Technica*, May 17. Available at: https://arstechnica.com/uncategorized/2007/05/google-book-search-adds-libraries-books-from-university-of-lausanne/ (Accessed: February 17, 2017).

Bartram, M. (2014) The History of e-books from 1930's "readies" to today's GPO e-book services. *Government Book Talk*, March 10. Available at: https://govbooktalk.gpo.gov/2014/03/10/the-history-of-ebooks-from-1930s-readies-to-todays-gpo-ebook-services/ (Accessed: May 19, 2017).

Bershidsky, L. (2017) Books stubbornly refused to be disrupted, and it worked. *Bloomberg View*, January 26. Available at: https://www.bloomberg.com/view/articles/2017-01-26/books-stubbornly-refused-to-be-disrupted-and-it-worked (Accessed: February 1, 2017).

Bloom, J. and Collins, R. (2015) Second Circuit holds that Google's digital copying acts are fair uses of copyrighted books. *The Licensing Journal*, 35(10), pp. 1-4.

Boote, D. and Beile, P. (2005) Scholars before researchers: on the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), pp. 3-15.

Bosman, J. (2011) Kindle users to be able to borrow library e-books. *The New York Times*, April 21, p. B4(L).

British Library (2011) *The British Library and Google to make 250,000 books available to all*. June 20. Available at: https://www.bl.uk/press-releases/2011/june/the-british-library-and-google-to-make-250000-books-available-to-all (Accessed: April 13, 2017).

British Library Board (n.d.) *Digitisation*. Available at: http://www.bl.uk/aboutus/stratpolprog/digi/digitisation/ (Accessed: April 17, 2017).

Brodeur, N. (2005) Deming's defender of words. *Seattle Times*. Available at http://www.seattletimes.com/seattle-news/demings-defender-of-words/ (Accessed: February 20, 2017).

Campon, J. (2007) The National Library of Catalonia joins the Library Project. *Google Books Search*, January 10. Available at: http://booksearch.blogspot.co.uk/2007/01/national-library-of-catalonia-joins.html (Accessed: February 17, 2017).

Carlson, S. (2006) U. of California will provide up to 3,000 books a day for Google to scan. *Chronicle of Higher Education*, 53(3), p. 32.

Center for Research Libraries (2005) *Principles and guidelines governing CRL agreements with commercial vendors for the digitization of holdings*. Available at: https://www.crl.edu/sites/default/files/attachments/pages/Guidelines%20for%20Vendor%20Digitization.pdf (Accessed: May 17, 2017).

Chandler, M. (2016) Google Books copyright suit ends as Supreme Court rejects challenge. *Investor's Business Daily*, April 18.

Chant, I. (2013) Lawsuit against Google Books dismissed. *Library Journal*, 138(20), p. 14.

Chartered Institute of Library and Information Professionals (2011) *User privacy in libraries*. Available at: http://www.cilip.org.uk/sites/default/files/documents/Privacy_June_AW.pdf (Accessed: February 20, 2017).

Chu, L. (2009) Google's book settlement is a ripoff for authors. *Wall Street Journal*, March 28, p. A9.

Clements, M. (2009) The secret of Google's book scanning machine revealed. *National Public Radio*, April 30. Available at: http://www.npr.org/sections/library/2009/04/the_granting_of_patent_7508978.html (Accessed: February 22, 2017).

Cliatt, C. (2007) Library joins Google project to make books available online. *News at Princeton*, February 5. Available at: http://www.princeton.edu/main/news/archive/S16/84/71S02/index.xml (Accessed: February 17, 2017).

Cohen, N. (2009) Millions of books, but no card catalog. *The New York Times*, February 2, p. B6.

Cole, L. (2014) BiblioTech: a library without books. *Chartered Institute of Library and Information Professionals*, August 12. Available at: http://www.cilip.org.uk/blog/bibliotech-library-without-books (Accessed: May 17, 2017).

Colombet, P. (2008) A first for France: the City of Lyon and Google partner up to digitize books. *Google Books Search*, July 22. Available at: http://booksearch.blogspot.co.uk/2008/07/first-for-france-city-of-lyon-and.html xml (Accessed: February 17, 2017).

Colombet, P. (2010) Google Books goes Dutch. *Google Books Search*, July 14. Available at: http://booksearch.blogspot.co.uk/2010/07/google-books-goes-dutch.html (Accessed: February 17, 2017).

Columbia University Libraries (2007) Columbia University Libraries becomes newest partner in Google Book Search Library Project. December 13. Available at: http://library.columbia.edu/news/libraries/2007/20071213_google.html (Accessed: February 17, 2017).

Conway, P. (2010) Preservation in the age of Google: digitization, digital preservation, and dilemmas. *The Library Quarterly: Information, Community, Policy*, 80(1), pp. 61-79.

Council of Europe (1988) Protocol to the convention for the protection of human rights and fundamental freedoms (European Convention on Human Rights). *Council of Europe Treaty Series 155*. Strasbourg: Council of Europe.

Digital Public Library of America (2015) *Public library partnerships project*. Available at: https://dp.la/info/about/projects/public-library-partnerships/selecting-content-for-a-digitization-project/ (Accessed: May 18, 2017).

Digital Public Library of America (2017a) *About*. Available at: https://dp.la/info/ (Accessed: May 18, 2017).

Digital Public Library of America (2017b) *History*. Available at: https://dp.la/info/about/history/ (Accessed: March 20, 2017).

Europeana (2017) *Europeana collections*. Available at: http://www.europeana.eu/portal/en (Accessed: March 20, 2017).

Goldsmith, K. (2013) The artful accidents of Google Books. *The New Yorker*, December 4. Available at: http://www.newyorker.com/books/page-turner/the-artful-accidents-of-google-books (Accessed: February 14, 2017).

Google Inc. (2004) *Google checks out library books*. Available at: http://googlepress.blogspot.co.uk/2004/12/google-checks-out-library-books.html (Accessed: February 2, 2017).

Google Inc. (2007a) *Cooperative agreement*. Available at: http://www.lib.utexas.edu/sites/default/files/google/utexas_google_agreement.pdf (Accessed: May 18, 2017).

Google Inc. (2007b) *Cooperative agreement*. Available at: http://www.btaa.org/docs/default-source/library/cic-googleagreement.pdf?sfvrsn=0 (Accessed: May 19, 2017).

Google Inc. (2011) *History of Google Books*. Available at: https://books.google.com/intl/am/googlebooks/history.html (Accessed: February 17, 2017).

Google Inc. (2016) *Privacy policy*. Available at: https://static.googleusercontent.com/media/www.google.com/en//intl/en/policies/privacy/google_privacy_policy_en.pdf (Accessed: February 14, 2017).

Google Inc. (2017a) *Books help*. Available at: https://support.google.com/books/partner/faq/3396243?visit_id=1-636216537189040585-3256143663&hl=en&rd=1#3474182 (Accessed: February 2, 2017).

Google Inc. (2017b) *Library partners*. Available at: https://books.google.com/googlebooks/library/partners.html (Accessed: February 2, 2017).

Google Ireland Limited (2011) *Cooperative agreement*. Available at: https://www.openrightsgroup.org/assets/files/pdfs/BL%20Google%20Contract.pdf (Accessed: March 25, 2017).

Green, A. (2010) Big digitisation: origins, progress and prospects. *International Journal of Humanities and Arts Computing*, 4(1-2), pp. 55-66.

Hall, K. (2011) Case study: digitising the British Library. *Computer Weekly*. Available at: http://www.computerweekly.com/feature/Case-Study-Digitising-the-British-Library (Accessed: May 19, 2017).

HathiTrust (2017) *Our partnership*. Available at: https://www.hathitrust.org/partnership (Accessed: March 20, 2017).

Hoffelder, N. (2013) Internet Archive now hosts 4.4 million eBooks, sees 15 million eBooks downloaded each month. *The Digital Reader*, July 9. Available at: http://the-digital-reader.com/2013/07/09/internet-archive-now-hosts-4-4-million-ebooks-sees-15-million-ebooks-downloaded-each-month/ (Accessed: March 21, 2017).

Hoffmann, A. and Bloom, R. (2016) Digitizing books, obscuring women's work: Google Books, librarians, and ideologies of access. *Ada: A Journal of Gender, New Media, and Technology*, 9. Available at: http://adanewmedia.org/2016/05/issue9-hoffmann-and-bloom/ (Accessed: February 14, 2017).

Howard, J. (2012) Google begins to scale back its scanning of books from university libraries. *The Chronicle of Higher Education*, 58(28).

Inouye, A. (2016) What's in store for ebooks? *American Libraries*, 47(1/2), pp. 42-47.

Internet Archive (2017) *Homepage*. Available at: https://archive.org/ (Accessed: March 21, 2017).

James, R. (2010) An assessment of the legibility of Google Books. *Journal of Access Services*, 7(4), pp. 223-228.

James, R. and Weiss, A. (2012) An assessment of Google Books' metadata. *Journal of Library Metadata*, 12(1), pp. 15-22.

Jaschik, S. (2007) An alternative to Google. *Inside Higher Ed*, June 22. Available at: https://www.insidehighered.com/news/2007/06/22/digitize (Accessed: February 24, 2017).

Jiji Press English News Service (2007) Keio Univ. to put 120,000 books online in tie-up with Google. July 6, p. 1.

Kahle, B. (2017) Transforming our libraries from analog to digital: a 2020 vision. *Educause Review*, March/April, pp. 26-36.

Kaufman, P. and Ubois, J. (2007) Good terms - improving commercial-noncommercial partnerships for mass digitization. *D-Lib Magazine*, 13(11/12). Available at: http://dlib.org/dlib/november07/kaufman/11kaufman.html (Accessed: February 26, 2017).

Kelley, M. (2011) Europeana libraries project begins digital aggregation. *Library Journal*, 136(16), p. 18(2).

Laing, A. (2006) Copyright in the age of mass digitisation. *Entertainment Law Review*, 17(5), pp. 133-138.

Leetaru, K. (2008) Mass book digitization: the deeper story of Google Books and the Open Content Alliance. *First Monday*, 13(10). Available at: http://journals.uic.edu/ojs/index.php/fm/article/view/2101/2037 (Accessed: February 22, 2017).

Library of Congress (n.d.) *Principles for Library of Congress third-party digitization agreements*. Available at: https://www.loc.gov/about/doing-business-with-the-library/third-party-digitization-agreements/ (Accessed: April 1, 2017).

M2 Best Books (2007) Bavarian State Library joins Google Book Search Library Project. March 9.

Macrina, A. and Glaser, A. (2014) *Radical librarianship: how ninja librarians are ensuring patrons' electronic privacy*. Available at: http://boingboing.net/2014/09/13/radical-librarianship-how-nin.html (Accessed: February 20, 2017).

Madison Public Library (2016) *USA Patriot Act and Madison Public Library records*. Available at: http://www.madisonpubliclibrary.org/policies/usa-patriot-act-and-madison-public-library-records (Accessed: February 20, 2017).

Manley, L. and Holley, R. (2012) History of the e-book: the changing face of books. *Technical Services Quarterly*, 29(4), pp. 292-311.

Mattiuzzo, G. (2010) A digital renaissance: partnering with the Italian Ministry of Cultural Heritage. *Google Europe Blog*, March 10. Available at: https://europe.googleblog.com/2010/03/renaissance-europes-period-of-cultural.html (Accessed: February 17, 2017).

McMillan, R. (2009) Privacy missing from Google Books settlement. *PC World*, August 28. Available at: http://www.pcworld.com/article/171089/article.html (Accessed: February 14, 2017).

Miller, L. (2010) The trouble with Google Books. *Salon*, September 9. Available at: http://www.salon.com/2010/09/09/google_books/ (Accessed: February 25, 2017).

Milliot, J. (2015) New study finds low levels of digital library borrowing. *Publishers Weekly*, 262(48), p. 4.

Milliot, J. (2016) As e-book sales decline, digital fatigue grows. *Publishers Weekly*, 263(25), p. 4.

Morgan, K. and Proffitt, M. (2017) *Advancing the national digital platform: the state of digitization in US public and state libraries*. Dublin, Ohio: OCLC Research. Available at: http://www.oclc.org/content/dam/research/publications/2017/oclcresearch-advancing-the-national-digital-platform-2017.pdf (Accessed: February 17, 2017).

Murrell, M. (2016) Unpacking Google's library. *Limn*, 6. Available at: http://limn.it/unpacking-googles-library/ (Accessed: February 24, 2017).

Musto, R. (2009) Google Books mutilates the printed past. *Chronicle of Higher Education*, 55(39), p. B4.

New York Public Library (2017) *Digital files for editorial, commercial, and other uses*. Available at: https://www.nypl.org/help/get-what-you-need/photographic-services/obtaining-images (Accessed: May 19, 2017).

Newman, J. (2011) The Google Books settlement: a private contract in the absence of adequate copyright law. *Scholarly and Research Communication*, 2(1). Available at: http://src-online.ca/index.php/src/article/view/29/47 (Accessed: February 15, 2017).

Nunberg, G. (2009) Google's Book Search: a disaster for scholars. *The Chronicle of Higher Education*, August 31. Available at: http://www.chronicle.com/article/Googles-Book-Search-A/48245 (Accessed: February 2, 2017).

Oxford English Dictionary (2001) 3rd ed. Oxford: Oxford University Press.

Prose, F. (2016) When librarians are silenced. *New York Review of Books*. Available at: http://www.nybooks.com/daily/2016/10/14/kansas-city-librarian-arrest-for-defending-free-speech/ (Accessed: February 20, 2017).

Reisz, M. (2009) Major errors prompt questions over Google Book Search's scholarly value. *Times Higher Education*, September 10. Available at: https://www.timeshighereducation.com/news/major-errors-prompt-questions-over-google-book-searchs-scholarly-value/408085.article (Accessed: February 2, 2017).

Research Information (2007) *Cornell University joins Google Book search*. August 10. Available at: https://www.researchinformation.info/news/cornell-university-joins-google-book-search (Accessed: February 27, 2017).

Rosenberg, S. (2017) How Google Book search got lost. *Backchannel*, April 11. Available at: https://backchannel.com/how-google-book-search-got-lost-c2d2cf77121d (Accessed: May 18, 2017).

Ruiz, J. (2011) Access to the agreement between Google Books and the British Library. *Open Rights Group*, August 24. Available at: https://www.openrightsgroup.org/blog/2011/access-to-the-agreement-between-google-books-and-the-british-library (Accessed: February 6, 2017).

Schmidt, E. (2005) Books of revelation. *Wall Street Journal*, October 18, p. A.18.

Simonson, M. (2015) E-books: the future? *Distance Learning*, 2015, 12(1), pp. 49-50.

Spacey, R., Cooke, L., Muir, A., and Creaser, C. (2014) Regulating use of the internet in public libraries: a review. *Journal of Documentation*, 70(3), pp. 478-497.

Thomas, C. (2010) What role should substantive fairness have in the English law of contract? An overview of the law. *Cambridge Student Law Review*, 6(1), pp. 177-196.

Thompson, C. (2013) Read free or die: Harvard hosts a nonprofit alternative to Google Books. *IEEE Spectrum*, 50(1), pp. 52-53.

Townsend, R. (2007) Google Books: is it good for history? *Perspectives on History*. Available at: https://www.historians.org/publications-and-directories/perspectives-on-history/september-2007/google-books-is-it-good-for-history (Accessed: May 17, 2017).

The University of Texas at Austin (2007) The University of Texas Libraries partner with Google to digitize books. *University of Texas Libraries News Release*, January 19. Available at: http://www.lib.utexas.edu/about/news/university-texas-libraries-partner-google-digitize-books (Accessed: February 17, 2017).

University of Wisconsin (2006) UW-Madison joins massive Google Book project. *University of Wisconsin-Madison News*, October 12. Available at: http://news.wisc.edu/uw-madison-joins-massive-google-book-project/ (Accessed: February 17, 2017).

Vaccaro, A. (2014) *Why it's difficult for your library to lend ebooks*. June 27. Available at: https://www.boston.com/news/technology/2014/06/27/why-its-difficult-for-your-library-to-lend-ebooks (Accessed: January 24, 2017).

Vasileiou, M. and Rowley, J. (2011) Marketing and promotion of e-books in academic libraries. *Journal of Documentation*, 67(4), pp. 624-643.

Weisl, A. (2010) Unlocking our shared cultural heritage: Google partners with the Austrian National Library. *Google Europe Blog*, June 15. Available at: https://europe.googleblog.com/2010/06/unlocking-our-shared-cultural-heritage_4958.html (Accessed: May 17, 2017).

Woo, S. (2006) Google's book project gets its first non-English partner. *Chronicle of Higher Education*, 53(7), p. A29.

Woo, S. and Trachtenberg, J. (2011) Amazon's Kindle will offer e-books from libraries. *Wall Street Journal*, April 21, p. B.4.

Wood, C. (2006) U.Va. library joins the Google Books Library Project. *UVA Today*, November 14. Available at: https://news.virginia.edu/content/uva-library-joins-google-books-library-project (Accessed: February 17, 2017).

Young, J. (2005) Authors' group sues Google over library-scanning project. *The Chronicle of Higher Education*, 52(6).

Zorn, T. and Campbell, N. (2006) Improving the writing of literature reviews through a literature integration exercise. *Business Communication Quarterly*, 69(2), pp. 172-183.