

An Overview of the Google Books Project and Other Digitization Initiatives: Implications for Libraries

Google Books 及其他數位化專案對圖書館的蘊義

Rosemary Spellman

Library Assistant II, Libraries Service Center,

Johns Hopkins University, U.S.A.

Email: rspellm3@jhu.edu

Robert P. Holley

Professor, School of Library and Information Science,

Wayne State University, U.S.A.

Email: aa3805@wayne.edu

Keywords (關鍵詞) : Google Books ; Collection Development (館藏發展) ; Digitization Projects (數位化專案) ; Digital Resources (數位化資源)

【Abstract】

This paper analyzes the effects of Google Books upon collection development efforts in academic libraries. Other mass digitization projects are also briefly discussed. The focus is on the hurdles that these projects face such as copyright questions, long-term technological planning, and continuing permanent funding. The authors also examine the role of digitization projects as a form of preservation and/or conservation. Finally, the paper looks at the culture of libraries, how use of library materials may change as a result of these projects, gives some ideas of how the public will want to access information, and presents some

thoughts about how libraries can maintain their relevance even as the public has access to a wealth of digital resources.

【摘要】

本文分析了 Google Books 對學術圖書館館藏發展的影響，並概述了其他大規模數位化的專案。本文重點在於對這些數位化專案所面臨的障礙之探討，如版權問題、長程技術規劃、和持久性資金撥款。作者也探討了這些數位化專案對圖書館館藏保存和保護所扮演的角色。文章最後審視了這些專案將如何改變公眾對圖書館資源的使用和獲取資訊的需求，並提供圖書館在面臨大量開放性豐富數位資源的情況下，如何保持其職能相關性之幾點思考。

Introduction

Google Books has certainly captured the public's attention, especially considering the number of lawsuits that have been brought against Google for various claims of copyright infringement. This paper gives an overview of this project and a few other mass digitization initiatives, discusses some of the technical and copyright issues, shows how digital archives can function both for access and for preservation, provides a brief overview of the culture surrounding libraries as well as that of the general public, and, finally, the evaluates these digitization projects with an emphasis on how Google Books has, and will continue to, affect collection development in large academic libraries. Overall, we think that Google Books will have a positive effect upon collection development but with some caveats that will be explained in this paper.

Several mass digitization initiatives are already underway. A few have made headlines while others are working in the shadows. Many are freely available to the public or freely available at libraries. Some of the larger players are Google, the Internet Archive, and the Open Content Alliance. Other smaller projects, which often focus on more specific areas, are useful to their subscribers even if they do not have a big name sponsoring them.

Overview of Projects

All of this started in 1989 with the Library of Congress and its "American Memory Project," which attempted to see if there would be an audience for "intellectual content in digital form." (Fineberg, 2009). Obviously, time has shown that there is an audience and a market willing to pay for such materials, although free information is always preferred. Some smaller projects include one from the Library of Congress, called "Chronicling America," which is a "free, national, searchable database of historic American newspaper pages published between

1880-1992" (Fineberg, 2009). Wayne State University contributed a selection of film negatives that comprise the "Virtual Motor City" collection (Pruzin, 2010). A similar project at the British Library seeks "to digitize 40 million historic pages" from their collection of newspapers ("British Library", 2010). Other initiatives include Project Gutenberg, the Universal Library Project/Million Book Project, and the International Children's Digital Library (Lackie, 2008).

Larger projects include JSTOR, Project Muse and the Internet Archive, which not only digitizes books but also archives websites with the Wayback Machine. Two of the largest ventures are the Google Books and the Open Content Alliance, which has the backing of Microsoft and Yahoo. Some initiatives focus on access to content by working to make OPACs widely available with projects such as OAIster and Open Library (Hutton, 2008; Lackie, 2008). HathiTrust's goal is to be a major access point as well as acting as a digital archive for materials for these above mentioned projects to access (Willett, 2009).

The three major players to be discussed in this paper are Google Books, the Internet Archive, and the Open Content Alliance. Google, starting with only five libraries, has said that it would "cost \$10 [each] to scan the 15 million documents set for digitization, and that the process could take a decade or more" (Flagg, 2005a). The Internet Archive operates all over the United States with special scanning stations and allows users to submit works of their own to the Archive. The Open Content Alliance was created partially as challenge to Google. It has more international input and is also searchable across multiple search engines whereas Google Books is only available via Google (Crawford, 2006). The Open Content Alliance also "will scan and digitize only texts in the public domain, except where the copyright holder has expressly given permission" (Flagg, 2005b).

Google Books has been criticized for being Anglo-centric. The problem arose from the tentative

settlement that forced Google to digitize only items that were “either registered with the U. S. Copyright Office or published in the United Kingdom, Australia, or Canada” (Flagg, 2010). Whether or not Google will be able to work with foreign rights holders is something that will change the perception of Google Books as being Anglo-centric in the future. It is also important to note that Google is completely dependent on other sources, principally libraries, for material. The primary goal is providing access with preservation as a secondary consideration. (Dougherty, 2010). “Google has made its methods for Google Books very clear: this is a search service with some viewing of the context of the search items” (Coyle, 2006). Presumably, Google had hoped this would be enough to avoid litigation; but their “opt-out” system was problematic for copyright holders. Thus, Google has faced several cases that placed restrictions on the materials Google has scanned (and will scan in the future). This has meant that if a “book...is under copyright, a user will be presented with three snippets of text, a count of the number of times the search term appears in the volume and links to online booksellers and information about the nearest local library that carries a print version of the book” (Proskine, 2006).

As Litman (2010) says: “where attention lands, litigation often follows.” Google Books certainly has the financial and star power behind its name to have high impact, but the potential for monopoly has caused concerns that smaller projects might be overshadowed. This star power has even eclipsed some similar projects that have managed to stay out of the courts (and thus, mostly out of the news) such as the Open Content Alliance, which has equally larger supporters such as Microsoft but is scanning only material that is definitively out of copyright. The Internet Archive has made a name for itself with the Wayback Machine and also seeks to create another type of digital library with varying formats, such as PDFs or text files that users can choose from to view the materials they desire.

Technical and Copyright Issues

Beyond the two most challenging technological issues, long term planning and metadata, the various lawsuits that Google has faced create constraints as well. As the lawsuits have progressed, they have led to changes to Google Books including several name changes. At one point, Google was able to give libraries that participated in its service only “a single digital copy [that would] allow faculty and students to read, print, download, or otherwise use five pages for books not commercially available...but not for interlibrary loan or e-reserves” (Fialkoff, 2008). And this would be for books already owned physically by the library! Other legal constraints were also required, including the number of terminals that could be used at public libraries to access the database of books (Flagg, 2010).

With regards to copyright issues, some questions have been raised about Google’s scanning entire works even if showing only snippets to users because the full copy still exists on Google’s servers. While there have not been any major failures in Google’s security systems, the possibility still exists that someone could hack the system and steal the copyrighted materials (Millot & Albanese, 2004). Overall, copyright law has not caught up with today’s technology, especially since “the internet has redefined and displaced the concepts of ‘ownership’ and accessibility” (Shuler, 2001). Jeweler (2005) with the Congressional Research Service believes that, if the concept of fair use is interpreted broadly during the ongoing litigation, Google’s scanning may be lawful. Litman (2010) discusses in detail some of the problems that have become apparent with the current copyright system within the United States. She claims that intermediaries (the copyright holders, in most cases) are arguing that digital distribution causes creators to not want to create; but, at the same time, these intermediaries want to monetize the very distribution streams they rail against. Also, Grimmelmann (2009) notes that the court system

has made the decision about “online copying as a fair use when the purpose is to create a Web search engine.” Another perspective is that “controlling copying is less important than controlling access to a work” (Proskine, 2006).

In some cases, users are able to print out entire copies of public domain titles that are part of the book search service. The Internet Archive and Project Gutenberg also provide this service. One complaint about Google’s possible copyright infringement is that Google is gaining more users who will then click on more advertisements and therefore increase its profit. Copyright holders claim this is an infringement on their ability to make a profit on the original product. Google also decided to move forward with an “opt out” process for copyright holders but was “relying on bibliographic metadata from partner libraries to determine copyright issues.” However, one factor distinguishes Google from some other more “library-like” services in that at least one author claims that the “Google Library Project [now Google Books] does not replace the need for books; it merely indexes them” (Proskine, 2006). There have been complaints over the quality of Google’s indexing, but that may be partially due to the fact that they are allowing searching against documents created by optical character recognition (OCR) without subsequent error correction (Coyle, 2006).

With any digitization project, one challenge is planning for the future. Technology seems to change almost monthly with something new and better coming out to replace digital items that may not be truly obsolete. One concern has been with different versions of works, especially for those that are born digital. How does one determine if something is a first edition when changes can be made that overwrite the original document? In addition, there is the issue of dealing with changing data formats. Anyone who has had to deal with the switch to Microsoft Office 2010 understands the frustration of dealing with varying file

formats. Furthermore, some worry about integrity problems with the media that stores various electronic data, such as “evidence that some CD-ROMs are displaying signs of deterioration far earlier than had previously been anticipated” (Reidy, 2002). Without a detailed plan in place to deal with these sorts of issues, any digitization project or collection is headed for failure. Google presumably has multiple backups available to deal with any potential loss of digital content.

There is also the complication of standardizing materials across several platforms, operating systems, and whatever other software is used to access information. This problem affects not only metadata, which, if standardized, would allow for greater cross-library searching, but the actual data itself. The United States Government is attempting to develop “good procedures for quality assurance and quality control” (Manus, 2008), but that is only part of the problems faced by various mass digitization projects. Google has come under scrutiny for having problems with scan quality and the resulting OCR version of documents and also for the quality of its metadata (Dougherty, 2010). Pope and Holley (2011) also discuss some of the problems relating to incorrect metadata for Google Books and the inadequate responses that have been provided when Google has been questioned about these deficiencies. While Google wants to place blame for errors on the originating library or the producer of the bibliographic data, Google should also acknowledge that some of these errors are of its own making.

Fialkoff (2008) puts one of the biggest challenges in simple terms: there is “no guarantee that Google will support its book base indefinitely.” Another concern is about a particular digitization project becoming a monopoly since its owner could raise prices to the point where libraries would be unable to afford its services. While libraries are given digital copies of the works they provide for Google to scan, various

restrictions have been placed on the use of these “library digital copies” (Grogg & Ashmore, 2007; Fialkoff, 2008).

As an example of the issues that using Google Books might create for libraries, in 2001, Shuler presented a study of a possible future for the federal government depository model. While we feel that the depository model still has some value, it may change from having depositories of physical materials to having depositories of electronic materials. However, for such a sort of “electronic depository” truly to become a reality, libraries must be able to rely on stable Internet access and be able to educate users on how to use the various services and projects that are available. Since library budgets are, at best, remaining stable but, more often, being cut; having a large electronic collection may not be something that a library can afford. Kichuk (2010) thinks that higher costs and economic troubles are what is keeping “universal accessibility” at bay and “if a library cannot provide electronic content, patrons will go elsewhere”.

Some believe that, if large research libraries never weed and then participate in a mass digitization project, there will be an imbalance of access to outdated information before current information is found (Tennant, 2005). Another technical challenge is merely finding the desired information. While WorldCat has long been held up as sort of “one-stop shopping,” sometimes “digital collections were frequently represented...as a single record for an entire collection” (Hutton, 2008). Also, Willett (2009) says that “there is no central service or discovery tool for an end user to discover whether or not a particular book or journal has been digitized by one of these projects.” Legal challenges have certainly changed Google Books, but it is important to remember that technological issues also come into play for such a long-term project as this one. With a 27% “average net annual growth rate” (Kichuk, 2010) for electronic resources, all these issues require attention.

Electronic Collections and Storage

As more and more materials become available online, electronic collections are now considered to be more valuable than print versions of the same materials. When working with or creating an electronic collection, one important task is to plan for the long term future in the same way that that preservation was an important consideration for print materials. However, “the fact that increasingly more information is moving from paper to electronic form, and is correspondingly more accessible, does not obviate the need for a permanent record and ongoing availability” (Barnes, 1997). This is especially true when one considers that any digitization project could suddenly cease without warning and that those digitized materials might be lost. While the Internet Archive’s “Wayback Machine” archives Web sites, this data could be lost if the website is not “crawled” frequently enough or if the data is housed in a database-backed Website inaccessible to the Web crawlers as well as for other less common reasons (Szydowski, 2010). The Internet Archive’s FAQ section describes how people can exclude their site using a robots.txt file but also discusses that “dynamic” pages that use coding other than HTML can be improperly archived, if they are even able to be archived at all (Internet Archive, n.d.).

Digitization projects must have some form of stable funding because, if a project is only funded by a grant or from extra money in operational budgets, the likelihood of an electronic collection disappearing increases (Davis, 2009). Libraries that have disposed of paper copies of journals when a digital copy became available should be concerned about the potentially fragile future of digital copies because most producers make “no assurance of long-term availability of the journals” (Barnes, 1997). Libraries are inclined to weed print copies that have digital alternatives to make more room available in the library for new materials and for other uses such as increased study space (Adams, 2009). Another advantage of digital collections is that “digitization also enables the

re-unification of collections and items that were conceived together, but have become physically separated” (Brindley, 2009). Researchers, who may have had otherwise to travel to many different sites in order to consult needed research materials, especially benefit from such digitization projects.

While materials are increasingly being digitized and, in some cases, born digitally, it is important to remember that digital collections are not as permanent as people may think they are. Libraries should take this into account whenever considering projects like Google Books. If libraries decide to undertake their own digitization projects, they face the same challenges with regards to copyright and technical issues. Only long-term planning for digital preservation and careful analysis of the items to be weeded in favor of electronic copies will keep any sort of electronic collection, even one created by the library itself, in good working order and available for users in the distant future.

Digitization as Preservation

Some argue that digitization is a way of preserving materials for the future; but it may not be a good first line of defense, especially with the possibility of corruption of digital copies and the issue of technological format changes. On the other hand, digital copies are one option when planning for disasters since they can be easily stored in an off-site location and most often take up less space than their physical counterparts. Palmer (2008) suggests that having redundant copies and well thought out plans along with stable digitization policies and procedures will help assure the most consistent practice over time. Section 108 of the Copyright Act allows libraries to have more than one copy of an item, which is especially important since libraries often need to have “an ‘iron mountain’ copy, a master copy, and a use copy, with only the use copy accessible at any one time” (American Library Association, n.d.).

Nonetheless, digitization has positive aspects for preservation and conservation. Ceynowa (2009) states that “the digitization by Google thus serves the purpose of collection conservation [since] information contained in the threatened books can be saved.” The Library of Congress has adopted this strategy for some of its more brittle items (Swartz, 2007). Having digital copies is a good way to preserve fragile documents because these copies allow multiple users to access documents that would be too fragile to examine physically. Preservation, however, is a side benefit of many digital initiatives. Capell (2010) says that, while digitization can help preservation goals, “the driving factor behind most digitization efforts is ready remote availability.” Certainly, Google did not initiate Google Books primarily to help libraries preserve their collections; but libraries should not overlook this added benefit since many books are at or near the point where they can no longer be handled without damage. The digitization of documents that are considered to have “intrinsic value” is also helpful in preserving items that are important enough to maintain in whatever original format they happen to be. The National Archives and Records Administration has detailed information about the definition of documents having “intrinsic value” (National Archives and Records Administration, 1999). Coyle (2006) also points out that “the production of a preservation-quality copy is somewhat contrary to the desire to digitize whole libraries quickly and inexpensively with the least amount of human intervention.” While these two ideas are not completely at odds with one another, some documents may require greater care than others. In the end, to avoid future problems, the goals of any digitization project, be it for a few hundred or a few million items, must be clearly defined before a single page is scanned.

As with all digital projects, the problems of data formatting and redundancy must be dealt with in order

to create a solution that will withstand the challenges of time. Shuler (2001) says that, no matter what, electronic information “must be durable, authentic, have validity, and [be] permeable.” Adams (2009) suggests that libraries might even be able to take a more active role and become a “publishers” of digital materials, although this fails to take into account the potentially murky financial and legal situations that would inevitably occur. Technical issues will likely forever dog the heels of any efforts to make digitization a first-line of defense against the carelessness of users and the ravages time upon materials, but digitization clearly has some part to play in protecting fragile documents.

Culture of Libraries

In a world where digitization is becoming increasingly popular and where an increasing number of documents exist only in a digital form, libraries must adapt to this situation and be willing to work with users to create a viable future. What will be the role of libraries in this future? If libraries are merely warehouses, then they can become data farms with nothing but racks and racks of servers holding electronic copies of materials for access by users all over the world. If, on the other hand, “libraries exist to preserve society’s cultural artifacts and to provide access to them,” (Proskine, 2006) then a different approach must be taken to adapt to the increasing emphasis on electronic materials. There is also the potential for a compromise where libraries continue to serve as access points for materials held by others.

As one example, the Internet Archive considers Web pages to be an important cultural artifact and therefore keeps copies of previous versions for the future. Dye (2006) says that “the goal of a research library...is to secure, preserve, and archive knowledge--all of it, because...if we don’t preserve and protect it, no one else will.” Once again, Google has become part of this process with its book scanning though they claim to

want to be only an indexing service and not necessarily provide access to the materials. Google does provide access in the sense that they give links for users to find copies of the materials they see on their screens, even if only as a snippet. Thus, Google should be considered to be an important potential partner for any sort of future projects, especially for research libraries as they face budget challenges that may cause problems with providing access to materials, especially ones that require supervision while being used. If an electronic copy is readily available, users can access that copy whenever and wherever they so choose, which frees up staff time for other projects.

The availability of digitized Google books may lead to a “just in time versus just in case” change in how libraries provide access to books. Barnes (1997) states that with this change in thinking, one must separate the issues of “access” versus “ownership.” If users are able to download a copy of a particular book to their computer (as Google Books allows for some items as well as does the Internet Archive, among others), the question of ownership becomes a bit trickier. Is a digital copy enough to constitute “ownership” of an item? With the increasing usage of devices such as Kindles and Nooks to license e-books, obviously people are beginning to consider a digital copy to be good enough to be considered owned. When items are available from a different library, even ones across the world, access is something that becomes more important, especially for those who may be doing research in areas that require travel to where the particular item is physically available but the researcher cannot easily afford the trip. Another twist is that “evidence is mounting that any material that is not available in digital form does not get used” (Hahn, 2006). Students and researchers have come to expect materials to be available digitally. Library expenditures indicate this change. Martell (2007) shows that “expenditures for electronic serials increased 2,175 percent” between 1995 and 2004.

Libraries should pay attention to possible future changes from the increasing reliance on digital materials and develop strategies to adapt to these changes in ways consistent both with their mission and with the desire of users. Libraries must also be willing to fund these changes. Some materials are harder to make decisions about than others. Fineberg (2009) argues that “as a primary source for local history information, all newspapers...are worthy of retention and preservation by libraries and archives.” Unfortunately, newspapers are usually oversized and printed on paper that is not meant to withstand the ravages of time, even in stored upon proper conditions. A second problem with electronic materials is that they may have various restrictions on use such as Google Books materials under copyright. In some cases, licenses for items such as electronic journals do not allow electronic copies to be used for services such as interlibrary loan or course reserves. Willett (2009) states that “some recent research has indicated that interlibrary borrowing has dropped as a result of increased electronic availability [but] it is likely that interlibrary borrowing will continue to rise, because people will discover more sources of interest without having full access to them.” Google Books is helping make materials more visible, and then, more accessible with links to publishers, booksellers, and WorldCat on their Google Books’ pages.

Millot and Albanese (2004) make the bold statement that “when all is said and done, Google is going to be the biggest library.” Personally, we feel that Google is not necessarily attempting to become a library but instead wants to act as a service to allow people to find the information they seek more quickly. Google wants, however, to make money in doing so from advertisements as well as from “keep[ing] 37% of revenue from online book sales” (Flagg, 2008). Finding materials has become a challenge for users and reference staff because they must be able to “identify appropriate resources and effectively search this wealth of electronic

content” (Kichuk, 2010). Libraries must adopt strategies to deal with the potential problems of digital materials and not consider them as temporary annoyances. While “the printed book or any document printed on paper will survive for many reasons” (Reidy, 2002), libraries must understand the wants of their users and avoid minimally funding short-term projects when they should be working to put longer-term solutions in place.

With all the emphasis on electronic materials and the development of electronic collections along with the subsequent weeding of print copies, special collections may become what distinguishes one library from another. Rieger (2010) makes the following important points on this matter. Challenges of having special collections become the key collection for some libraries include users’ needing to learn several different interfaces to access the information they want, the lack of metadata standards, and the fact that some users still need to have access to the physical items even when the content is available digitally. Overall, libraries have not completely figured out how to deal with special collections in the digital age.

Finally, libraries must be willing and able to “add value beyond the search engine” and accept the fact that “documents are increasingly regarded as commodity services” (Brindley, 2009). Libraries must understand that some see them as outdated. We feel, however, that libraries could change this perception if they focused even more on concepts such as information literacy. While Google and Google Books may offer massive amounts of information in milliseconds, users can quickly be overwhelmed and get bogged down by information overload if they cannot distinguish between “good” or “bad” information. Libraries can remedy this situation through reference services and information literacy instruction for users. Of all the competitors in the information world today, libraries may be the best suited for this task. The snippets that Google Books provides for most copyrighted works are not enough for most users to judge the potential of the item to meet their needs.

A survey by Dunn and Menchaca (2009) had faculty and librarians rank the top three roles (purchaser, archive, information gateway) of libraries. The results showed a disconnect between users and librarians. Faculty thought that purchaser was most important role, but librarians believed information gateway was. However librarians feel about their role, they will need to become knowledgeable about the technologies that are important to their users. Basically, the function of libraries is changing in an increasingly technical-oriented society; they must adapt or be left behind as quaint relics that store artifacts that no one cares about any longer.

Culture of Users

The general public has increasingly global tastes and expects a free flow of information from anywhere in the world through the intermediary of the Web. Libraries must understand and accommodate this desire. The term “information society” includes the idea that the world is a lot smaller than it used to be and that people are now focusing on using information as a tool to achieve results beyond the mere satisfaction of fact based reference queries. As more scholarly materials appear freely on the web, such as lectures and other materials offered through iTunesU (<http://www.apple.com/education/itunes-u/>), autodidacts have even more resources available for self learning. Along with globalization, people expect materials to be available 24 hours a day, even if located halfway around the globe.

Shuler (2007) suggests that “users, publishers, and other participants in the global knowledge industry consider the library, as an institution, to be no longer a critical partner in shaping the evolving social relationships building on these digital environments.” We believe they may be half correct in this assessment. Right now, libraries are on the verge of being left behind; but we feel that many libraries have made

strides to be more active in these “evolving social relationships” by making use of various Web 2.0 creations such as YouTube videos, Facebook pages, and Twitter accounts. These resources, however, deal only with the user side of the equation. Information must also be as easily accessible as the tools that contain it. If Google is the path of least effort, then the public will use Google.

Shuler (2007) also has an interesting perspective on accessing information, namely that: “a nation's long-term survival depends on how well any of its citizens can access an increasingly complex global information structure.” Once again, if Google is the easiest option and provides acceptable, if not the best results, it will become the dominant resource. The general public does not particularly care about the source of the desired information; it just wants the information right now. And in some cases, that information may be incorrect or incomplete; without libraries to help judge its quality, it may do more harm than good. Tennant (2005) says that someone “who finds a pre-1923 source available for free via Google is unlikely to sashay down to the local library for something more recent.” In some cases, pre-1923 materials may be all that a user needs; but, most often, more recent information is more useful. Information overload has increasingly become a problem in contrast to information scarcity in pre-Internet days, but developing information literacy (and not just in a one class at the start of the freshman semester but continuing throughout a user's life) can combat the glut of information to some extent (Badke, 2010). Another issue is the “paradox of choice,” which claims that, as someone has more options for a particular information need past a certain tipping point, making a decision becomes impossible (Schwartz, 2004). Libraries should think about how they can make sure that the best materials are available for user needs. Providing easy access to pre-1923 materials rather than more recent items is not the optimal choice.

Libraries must also consider how people want to use the Web and the information they find there. Hutton (2008) says that students want library websites to be a portal where “everything they need [is] pulled together into a single place.” While some libraries may meet this demand, doing so requires developing standards for sharing materials among libraries. Google’s only allowing Google Books to appear in their searches exclusively works against the desire for universal access. Since users are viewing the Internet more as a series of portals rather than as discrete silos and also increasingly want to be involved in creating and publishing material not just consuming it, libraries will need to work with the general public to meet their expressed and implicit desires.

Value of These Projects

For a digitization project to be valuable, it must have a target audience that is willing to support it, either by using the collection at a high enough level to meet the producer’s expectations or monetarily through fees or donations. Various groups will have different opinions about the “value” of a particular product. Google provides something valuable for just about everyone. Milliot (2007) says that publishers are increasing their involvement in Google Books because they are getting revenue from the links on the pages of snippets by building their direct-sales business because the links lead to the publisher’s main page. In addition, publishers are finding that requests are for older titles have increased since “20% of the clicks are for books that are over 10 years old” (Milliot, 2007). In a similar way, Google’s providing links to library holdings definitely falls in line with a library goal of providing more access and may rejuvenate the use of older materials in the library. Another source of revenue of potential revenue may be the creation of smaller sets of digitized works for sale (Holley, 2009) on the model of the pre-Internet microform sets on specific topics or covering defined periods. Hahn (2006) believes that

mass digitization “means that nothing will ever go out of print.” The availability of digital Google Books may affect the sales on the out-of-print market. For the first five libraries involved in Google’s mass digitization project, “about 65% [of the items] are out of print” though many are still under copyright, which is a major issue for Google and those who wish to purchase these items.

Lackie (2008) has a more negative view of the situation and claims that there are “no real world examples of innovative ways that teachers and librarians are using [Google Books] to help students learn.” While it may be true that people are not using Google Books in ways that are particularly innovative, the increased availability of such a large number of materials provides value by supporting the traditional goals of teaching, research, and reading for pleasure. Another issue may be the difficulty for librarians in keeping up with the vast number of digitization projects to the point that they can instruct users on how to use any particular one. For researchers, “digitization allows for separation of content and carrier” (Tian & Martin, 2009) by providing greater access to items that may not be readily available wherever the researcher is located.

Effect of Google Books on Collection Development

Google Books has definitely had an effect on academic libraries, especially larger ones that are willing to pay for access to more esoteric texts. In addition, Google has other services that are useful to libraries as well, such as Google Scholar and Google Translate. Furthermore, since libraries are not immune to running out of space, Lackie (2005) is concerned that Google Books is “another instance of the JSTOR effect, where big libraries get rid of low-use items simply because they can now subscribe to databases that offer pretty good searchable likenesses of them.” Libraries should keep in mind the risks of the potential

loss of digital access when weeding based on what is currently available online. While we think that the situation is not quite as dire as some experts make it sound, we also understand the wariness about the permanence of these initiatives as has been discussed earlier in this article.

It is important for libraries to remember that nature abhors a vacuum. "If academic libraries offer insufficient access to free full-text-e-books, the Google search engine is well positioned to direct users to the content of its digital book project and perhaps additional scholarly full-text resources" (Hutton, 2008) even with the technical issues discussed above regarding copyright and dealing with varying data formats. While some caution against relying on commercial entities for access to materials (Hutton, 2008), more and more services are being offered that way. Libraries must learn to adapt. An increase of information available online has created an upward spiral of expectations. As information is found online, people expect even more things to be online and have higher expectations on the availability of resources than in the print age (Rieger, 2010). Personally, we feel that Google Books will have an overall positive impact on academic libraries; but this judgment does not negate the fact that some issues need to be addressed.

Google Books has become the most discussed mass digitization initiative. This attention is partially well deserved for the riskier moves that Google has made, especially concerning copyright infringement. There are a number of other projects available, but no one well-known source exists currently to search across all the projects to find if a particular item has indeed been digitized. Rieger (2008) discusses the "DLF/OCLC Registry of Digital Masters (RDM)" and how items that are registered there have been "created under established best practices for digitization and that the institution was committed to its digital preservation." Better ways to find out easily about other projects will

need to be developed for them to become mainstream for libraries. Another potential issue is that Google's contracts with libraries are non-exclusive (Grogg & Ashmore, 2007). Thus, materials may be duplicated among different projects. Having RDM or some other form of registry for digitized items widely accessible for library use may resolve this problem and also help libraries find out if something has been digitized and available for users. Libraries themselves must understand that their role is changing as technology becomes a greater force in today's society and that electronic collections are a reality that they should plan for now instead of waiting for a crisis to occur. As items that have shaped thinking around the world and have fostered research in many disciplines begin to age to the point where use creates the potential for permanent damage, digitization offers a chance for these items not to be locked away for protection but to be used by more users than ever imagined, especially if they are posted online for free by a service such as Google Books or the Internet Archive. To conclude, we think that Google Books is a positive force for libraries but that the technical and legal issues will need to be settled first. Libraries also need to plan ways to make this wealth of digital resources a mainstream source of knowledge that is appropriate for many types of users and that creates added value for them.

References

- Adams, W. R. (2009). Archiving digital materials: An Overview of the issues. *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve*, 19(4), 325-335.
- American Library Association (n.d.). *Library preservation: Changes incorporated in H. R. 2281 the digital millennium copyright act of 1998 (PL 105-304)*. Retrieved from <http://www.ala.org/ala/issuesadvocacy/copyright/dmca/pdfs/librarypres.pdf>
- Badke, W. (2010). Information overload? Maybe not. *Online*, 34(5), 52-54.

- Barnes, J. (1997). Electronic archives: An Essential element in complete electronic journals solutions. *Information Services & Use*, 17(1), 37-48.
- Bridley, L. J. (2009). Challenges for great libraries in the age of the digital native. *Information Services & Use*, 29, 3-12.
- British library, Brightsolid to digitize historic newspapers. (2010). *Advanced Technology Libraries*, 39(7), 3.
- Capell, L. (2010). Digitization as a preservation method for damaged acetate negatives: A Case story. *The American Archivist*, 73(1), 235-249.
- Ceynowa, K. (2009). Mass digitization for research and study: The Digitization strategy of the Bavarian State Library. *IFLA Journal*, 35(1), 17-24.
- Coyle, K. (2006). Mass digitization of books. *The Journal of Academic Librarianship*, 32(6), 641-645.
- Crawford, W. (2006) Building the econtent commons. *Econtent*, 29(2), 44.
- Davis, M. (2009). Finding the middle way in sustainable digitization efforts. *Against the Grain*, 21(5), 34-38.
- Dougherty, W. C. (2010). The Google Books project: Will it make libraries obsolete. *The Journal of Academic Librarianship*, 36(1), 86-89.
- Dunn, R. & Menchaca, F. (2009). The Present is another country: Academic libraries, learning technologies, and relevance. *Journal of Library Administration*, 49(5), 469-479.
- Dye, J. (2006). The Digital rights issues behind book digitization projects. *Econtent*, 29(1), 32-37.
- Fialkoff, F. (2008). Google deal or rip-off. *Library Journal*, 133(20), 8.
- Fineberg, G. (2009). One million pages: Digitized newspapers preserve nation's history. *Library of Congress Information Bulletin*, 68(7/8), 148-150.
- Flagg, G. (2005a). Google partners with libraries in massive digitization project. *American Libraries*, 36(1), 26-27.
- Flagg, G. (2005b). Yahoo, European Union announce digital library projects. *American Libraries*, 36(10), 22-23.
- Flagg, G. (2008). Google settles scanning suit with authors, publishers. *American Libraries*, 39(11), 30.
- Flagg, G. (2010). Revised Google Books settlement tackles foreign titles, orphans. *American Libraries*, 41(1/2), 30.
- Grimmelmann, J. (2009). Why it matters. *Publishers Weekly*, 256(47), 20-21.
- Grogg, J. E. & Ashmore, B. (2007). Google Book Search libraries and their digital copies: What now? *Searcher*, 15(4), 18-27.
- Hahn, T. B. (2006). Impacts of mass digitization projects on libraries and information policy. *Bulletin of the American Society for Information Science and Technology*, 33(1), 20-24.
- Holley, B. (2009). Three cheers for the Google Books project. *Against the Grain*, 21(4), 44.
- Hutton, J. (2008). Academic libraries as digital gateways: Linking students to their burgeoning wealth of open online collections. *Journal of Library Administration*, 48(3/4), 495-507.
- Internet Archive Frequently Asked Questions. (n.d.). *Internet archive*. Retrieved March 1, 2011 from <http://www.archive.org/about/faqs.php>
- Jeweler, R. (2005). *The Google Book Search: Is online indexing a fair use under copyright law?* Washington, DC: The Library of Congress, CRS-Congressional Research Service.
- Kichuk, D. (2010). Electronic collection growth: An Academic library case study. *Collection Building*, 29(2), 55-64.
- Lackie, R. J. (2005). Google's print and scholar initiatives: The Value of and impact on libraries and information centers. *Internet Reference Service Quarterly*, 10(3/4), 57-70.
- Lackie, R. J. (2008). From Google print to Google Book Search: The Controversial initiative and its impact on remarkable digitization projects. *The Reference Librarian*, 49(1), 35-53.

- Litman, J. (2010). Real copyright reform. *Iowa Law Review*, 96(1), 1-55.
- Manus, S. (2008). Federal guidelines: Agencies collaborate on digitization. *Library of Congress Information Bulletin*, 67(12), 218-219.
- Martell, C. (2007). The Elusive user: Changing use patterns in academic libraries 1995-2004. *College & Research Libraries*, 68(5), 435-444.
- Milliot, J. (2007). Google's Book Search moves forward. *Publishers Weekly*, 254(22), 14.
- Millot, J & Albanese, A. (2004). Google deal with libraries wins praise, raises questions. *Publishers Weekly*, 251(51), 7.
- National Archives and Records Administration. (1999). *Intrinsic value in archival material. Staff information paper number 21*. Washington, DC: National Archives and Records Administration.
- Palmer, G. (2008). Digitization, can it play a role in disaster preparedness: Notes from the field. *Oregon Library Association Quarterly*, 14(4), 20-22.
- Pope, J. T. & Holley, R. P. (2011). Google Book Search and metadata. *Cataloging & Classification Quarterly*, 49(1), 1-13.
- Proskine, E. A. (2006). Google's technicolor dreamcoat: A Copyright analysis of the Google Book Search library project. *Berkeley Technology Law Journal*, 21, 213-239.
- Pruzin, C. A. (2010). *State digital resources: Memory projects, online encyclopedias, historical & cultural materials collections*. Retrieved from <http://www.loc.gov/rr/program/bib/statememory>
- Reidy, D. V. (2002). The Electronic Icarus: Some problems and some solutions in digitisation. *Information Services and Use*, 22, 151-157.
- Rieger, O. Y. (2008). *Preservation in the age of large-scale digitization*. Washington, DC: Council on Library and Information Resources.
- Rieger, O. Y. (2010). Enduring access to special collections: Challenges and opportunities for large-scale digitization initiatives. *RBM*, 11(1), 11-22.
- Schwartz, B. (2004). *The Paradox of choice: Why more is less*. New York: Ecco.
- Shuler, J. (2007). Academic libraries and the global information society. *The Journal of Academic Librarianship*, 33(6), 710-713.
- Shuler, J. A. (2001). Beyond the depository library concept. *The Journal of Academic Librarianship*, 27(4), 299-301.
- Swartz, N. (2007). Library of Congress to digitize brittle books. *The Information Management Journal*, 41(3), 12.
- Szydlowski, N. (2010). Archiving the web: It's going to have to be a group effort. *The Serials Librarian*, 59(1), 35-39.
- Tennant, R. (2005). Google out of print. *Library Journal*, 130(3), 27.
- Tian, X. & Martin, B. (2009). Business models indigital publishing: Some insights from Australia. *Publishing Research Quarterly*, 25(2), 73-88
- Willett, P. (2009). Mass digitization and its impact on interlending and document supply. *Interlending & Document Supply*, 37(3), 143-148.