

Preservation Pending:

## The Future of E-book Access in the Digital Age

Melissa Goertzen  
Robert Wolven  
Jeffrey D. Carroll<sup>1</sup>

### Abstract

For a very long time human knowledge has been recorded on print media and passed down, or preserved, for future generations. Since the latter half of the Twentieth Century however this human knowledge has been increasingly recorded in digital media. Whether we will be as successful as our predecessors at preserving human knowledge will depend on the steps we take now to identify and address the risks and threats when knowledge is stored in digital form. This report will focus specifically on e-books as a subset of all digital media. In some respects we are lucky in that the conversation surrounding the risks and threats to journal literature as it passed from print to digital has been going on for more than two decades, and some promising models and ideas have emerged. On the other hand e-books present problems that differ from those on the journal side and which will need to be addressed from scratch, as it were, with no prior models on which to build. Nor will this report provide ready answers. Rather, by bringing the conversation here we are throwing down the gauntlet and challenging all members of the information ecosystem to think seriously about the issues and to take the steps necessary to work toward a solution that is of mutual benefit to us all.

### Introduction

The term preservation in libraries refers to the overall system of efforts aimed at maximizing the longevity of our cultural heritage and of the outputs of scholarly inquiry. Research is a cumulative activity; scholars of today build on the findings and hypotheses of scholars that have gone before. As scholarly interests and societal values change, works that may have seemed to have limited, ephemeral interest when first published may take on new importance. Libraries have been the primary means of ensuring the survival of the scholarly and cultural record. For printed materials, they have accomplished this through *conservation*—the treating of single items in an effort to mitigate the effects of deterioration— by storing materials in controlled environments, and by converting the content to new formats when necessary. Each of these actions is but one element of a preservation strategy. Preservation practices evolve in the delicate balance between art and science. With each new generation, volumes of intellectual and cultural output undergo rigorous evaluations to determine which artifacts merit preservation efforts. Conservation efforts are made to preserve both the item itself as well as its content.

---

<sup>1</sup> Melissa Goertzen is the E-Book Program Development librarian at Columbia University Libraries. Her research focuses on the documentation of the e-book landscape and the development of strategic plans related to e-book collection development programs. Robert Wolven is associate university librarian for bibliographic services and collection development at Columbia University Libraries. He is a former co-chair of the American Library Association's Digital Content and Libraries Working Group (DCWG) and a current member of the DCWG subgroup on digital preservation. Jeffrey D. Carroll is director of Collection Development at Columbia University Libraries. His work most recently has focused on the opportunities and challenges of cooperative collection development and vendor negotiations across multiple research institutions.

Then, efforts are made to place items into a historic context that forms the basis for a common cultural heritage. However, as society becomes increasingly reliant on digital technologies to produce and disseminate information, the development of strategies to preserve this information has lagged behind. This is due in large part to the rate of technological changes taking place in the sharing of information and the complexity involved in constructing an environment that accounts for the quick pace at which this rapidly evolving technology becomes obsolete. The preservation of digital content becomes a daunting task facing information professionals.

Today, much discussion surrounds the development of e-book preservation practices. In many cases, proposed strategies to tackle this complex challenge are still theoretical and abstract. However, last year librarians at Columbia University Libraries (CUL) received an acquisition request for a unique born digital item that rooted our preservation challenges in an entirely practical context.

In 1963 Josef Albers published the *Interaction of Color*, which is hailed as the canonical work regarding the study of and relationships between colors. Albers developed the text to serve as a hands-on kit that provides an interactive environment to conduct silkscreen color studies. However, there are limits to interactivity as the book was originally published in print format. Now, fifty years after its original publication the text has been released in a digital format that brings Albers' original vision to life. Yale University Press released the *Interaction of Color* app, which includes over 125 color plates, 60 interactive studies, and commentary from Albers. This is a landmark publication and won the George Wittenborn Memorial Book Award for excellence in art publishing from the Art Libraries Society of North America in the past year (YaleNews, 2014).

Librarians at CUL faced a number of central challenges. Because its value to the scholarly community is tied to its interactive learning environment, we searched for a means to purchase a preservation copy that guarantees long-term access to both content and functionality. We discovered that because the app can only be licensed through iTunes (currently, there is not a business model that supports a flat out purchase of the app), there is no way to capture and archive content. Even if there were such a method, there are no benchmarks that estimate the costs of long-term app maintenance and storage.

While there is debate regarding whether this item should be considered an e-book, it still raises valid questions about e-book preservation practices and strategies. Until recently, e-versions of texts have been viewed as supplementary to print copies (Armstrong and Lonsdale, 2009). Since they were acquired to "back up" print copies, preservation of the e-version was not considered urgent or even necessary. Today, we stand at a tipping point and discover an increasing number of e-books that have no print equivalent. Like the *Interaction of Color* app, many of these items have significant cultural and intellectual value that may impact the way knowledge is disseminated within the research community and general public in the coming years.

In today's information society, securing the funds and server space to preserve text-only files can be a challenge. However, this process seems straightforward compared to what e-books are becoming: interactive pieces of media that are hybrids of databases, websites, and games (Robertson, 2014). In order to develop effective preservation strategies and frameworks, the library community first must understand on an intellectual, technical, legal, and financial level what preservation means for items in the digital world (Robertson, 2014). For instance, does digital preservation apply to content, functionality, or both? Is it possible to secure legal rights to preserve content that is licensed? How do we guarantee perpetual

access to content that is tied to hardware, software, and file formats that will become obsolete within the next decade?

While the answers to the above questions will shape future preservation policies, a larger challenge looms on the horizon – the sheer size of the digital universe. With an exponential growth in born digital content, there is a massive body of electronic information at our fingertips – portions of it merit preservation while others do not. In this environment, how do librarians decide what items should be captured, stored, and managed over time? The answer will impact how our intellectual and cultural heritage is shaped for future generations of scholars, students, and readers.

### **The Landscape: E-books vs. E-journals**

By 2005, the information profession faced a significant digital preservation challenge with the growth of e-journal publication and many libraries beginning to collect some journals solely in digital form. At that time, the academic community recognized a pressing need to take action to preserve scholarly electronic journals to guarantee perpetual access to information. The need for a stable and reliable preservation strategy was outlined in a report entitled *Urgent Action Needed to Preserve Scholarly Electronic Journals*, which was authored by library leaders representing research institutions in the United States and received the support of the Association of Research Libraries (ARL) (Kirchhoff, 2011). The document defined the initial strategy as a “way of managing risk, first, against the permanent loss of electronic journals and second, against having journal access disrupted for a protracted period following a publisher failure” (ARL, 2005). In the following years, a task force under the guidance of OCLC’s Research Libraries Group (RLG) and the National Archives and Records Administration (NARA) developed standardized tools and metrics. The most notable was the Trusted Repository Audit Checklist (TRAC) based upon the Open Archival Information Systems (OAIS) Reference Model. Today, digital repositories that have passed TRAC certification have the highest level of protection and security of e-journal content. Examples of these repositories include CLOCKSS (Controlled Lots of Copies Keeps Stuff Safe), Portico, and HathiTrust (Center for Research Libraries, n.d.).

Given the fact that standardized metrics now exist for e-journal preservation (and to date, have proven successful), one may ask why TRAC has not been more widely applied to e-book preservation. After all, both formats share similarities in their general makeup – they are both electronic resources that consist of two parts: files and metadata. However, when the surface is scratched, e-book preservation presents unique challenges that have not been encountered in the past. These include, but are not limited to, the following six issues:

1. E-book versions – journal articles, once published, tend to be stable; by contrast e-books – particularly trade books – may continue to be revised and enhanced;
2. Digital rights management (DRM) – relates to the common practice of embedding technologies in e-book files or devices that restrict access to content based on licensing terms;
3. Metadata – relates to managing the hierarchy of metadata at many levels of the publication (e.g. chapters, volumes, and series);
4. Legal issues – refers to the rights to use, share, and preserve e-books;
5. Format issues – refers to the coupling of e-book content with hardware, software, or distribution platforms;

6. Business models – refers to the sustainability of e-book licensing models and a fractured business landscape driven by proprietary platforms.  
(Kirchhoff & Morrissey, 2014), (Kelley, 2014), (Kirchhoff, 2011).

With this complex e-book landscape, libraries must develop preservation practices and strategies that promote perpetual accessibility, usability, authenticity, and discoverability of information. Together, these four qualities ensure “fitness of use,” a primary principle that guides and informs ongoing preservation decisions, actions, and financial investments (Smith, 2004).

## **Global Trends**

This report primarily focuses on preservation trends and challenges faced by stakeholders in the United States. However, as markets become increasingly global in nature, it is important to understand e-book landscapes outside of North America as they will impact preservation strategies in the future – particularly in terms of legal rights and collection development priorities.

In the English-speaking world, the U.S. has the strongest e-book market – e-book sales account for approximately a third of book revenues, particularly in the segments of adult fiction and genre fiction (Rüdiger Wischenbart, 2014). Across Western Europe, the adoption of e-books by consumers varies. In Germany, which has one of the largest book markets in the world, e-book purchasing trends are beginning to follow English language markets, but are approximately three years behind (Rüdiger Wischenbart, 2014). Across France, Spain, and Italy, e-book “momentum seems to be fairly limited” (Rüdiger Wischenbart, 2014, p. 20). Most surprising are trends observed in Sweden, where a “domestic e-book market has hardly taken shape, due to a mix of high prices for e-books and the role taken up by libraries who are lending significantly more e-books to readers than retail has sold so far” (Rüdiger Wischenbart, 2014, p. 20).

In other parts of the globe, a number of large markets are emerging. Below are four markets that may impact the future of online educational resources.

1. China: The world’s most populous country is also the second largest e-book market in the world, with 52 percent growth in 2012 (Kelbanoff, 2013). It is also estimated that there are 500 million Internet users and 1.22 billion mobile phone users in the country (Kelbanoff, 2013). Based on the market potential, competition among e-book providers is growing. In 2013, Amazon launched a dedicated Chinese platform, Apple’s iPhones were cleared to run on China’s mobile networks, and the domestic online platform, Dangdang, offered its catalog for free (Rüdiger Wischenbart, 2014). Currently, the fastest developing market for English-language books is focused on libraries, schools, and other institutions (Kelbanoff, 2013).
2. Russia: During 2013, Russian e-book sales surpassed those in the United Kingdom and Brazil, making Russia the third largest e-book market in the world (Gerden, 2014). The country has a “thriving reading culture in which writers and intellectuals occupy a prominent role in the public sphere and in which books stand at the center of the country’s cultural ambitions” (Rüdiger Wischenbart, 2014, p. 58). Since 2011, the market increased by nearly 200% and is expected to continue (Gerden, 2014). Although the sale of dedicated e-readers has declined, sales from

personal computers and laptops continue to drive the market forward. Since 2012, Apple opened an iTunes store in Russia, and Google opened a Play store, and Kobo announced plans to offer services (Rüdiger Wischenbart, 2014).

3. Brazil: The Brazilian book market has shown strong growth for years with no signs of slowing down. For instance, consumers in 2013 spent 8.8 billion Brazilian reais (over \$4 billion) in books, a growth of 7% from the previous year (Utsumi, 2014). Currently, e-books are not widely used, but companies like Amazon, Apple, Google, and Kobo (all of whom set up e-book stores in the country within the past two years) believe there is large potential for development. In 2013, a number of Brazilian publishers attributed 4% of their market shares to e-books (Utsumi, 2014). Another factor that is expected to influence the market is the fact that customers do not pay value-added taxes or sales taxes on books in Brazil (Rüdiger Wischenbart, 2014). “The Brazilian senate is discussing a bill that would officially make both e-books and dedicated e-readers tax-free. If the bill passes, the prices of E-Ink Kindles, Kobo devices, and Nooks would have to come down steadily” (Rüdiger Wischenbart, 2014, p. 65).
4. India: Currently, close to 24% of texts published in India are English-language titles. To date “all the major publishers in India, such as Penguin Books India, Hachette India, and Westland, have digitized their English-language backlists” (Vyas, 2014, para. 2). There is also great progress taking place in terms of regional e-books. With 25 regional languages spoken across the country, there is an opportunity for writers “outside of the country to make further inroads into the Indian market through getting their works translated” (Vyas, 2014, para. 4). This is especially true for self-published titles, as the legal process of forming partnerships with translation companies is less complex. On the technological side, dedicated e-readers have not made a significant market impact due to consumers’ questions about their overall value, but companies like Flipkart have successfully introduced e-reading apps that can be used on tablets and phones, allowing consumers to “read e-books on devices they are most familiar with” (Vyas, 2014, para. 9).

These global trends are likely to prove of increasing importance to libraries in the U.S. as well. Many public libraries acquire books from other countries to serve foreign-speaking populations. Research libraries acquire a range of publications from all world regions to serve both current and future scholarship. Just as e-book markets are developing in different ways and at different rates, though, the legal regimes, licensing terms, and business models that affect the ability of libraries to preserve this content present wide differences, and thus complicate the challenge.

### **Current Preservation Trends**

The Library Journal’s e-book survey discovered a 93 percent increase in e-book collections among academic libraries since 2012. The survey also found that libraries anticipate e-book spending to comprise 20 percent of their budgets within five years (Blummer & Kenton, 2012), a figure that could range anywhere from a half million dollars for a medium-sized academic library to over two million for a large one. Despite the increase in e-book acquisition, it is not clear if or how e-book content will remain accessible in years to come.

Historically, the first-sale doctrine has provided a crucial underpinning for preservation work: when libraries purchase a print book, they are free to rent it out, resell it, or keep it forever (Smith, 2004). In this environment, libraries own their copy of the physical book, own the storage unit (often a bookshelf), and maintain the book for future use. Digital objects have disrupted this model. The idea of purchasing a “copy” of an e-book or “lending” it to someone else is no longer relevant. In the digital world, the first-sale doctrine is no longer applicable in a legal context. In 2010, an appeals court ruling in the case *Vernor v. Autodesk* established that individuals cannot “buy” a piece of software, only acquire a permanent and non-transferable license to it (Smith, 2004). For libraries, this means that the legal ownership of individual titles, the storage unit (often a piece of hardware or software), and the ability to maintain files for future use are tied to the content provider – often a publisher or software developer.

Libraries license e-books and usually these licenses do not include the legal right to capture or preserve content (Billington, 2013). In the past, the inability to preserve e-book content has not been a concern as long as a print version is available in the collection. However, as born digital content becomes increasingly common, the information profession faces a renewed urgency to preserve content that does not have a print equivalent. This is particularly problematic when DRM restrictions prevent libraries from downloading or printing copies of e-books for archival purposes. In the current market, publishers are often not prepared to sell digital e-book files to libraries due to fears of piracy, decreased revenue, or an absence of business models that support such transactions.

On the flip side, even if publishers were prepared to sell e-books files, the majority of libraries do not have adequate infrastructure to house them. Though there are some notable exceptions, at this time, most do not have an institutional repository capable of storing digital files for generations to come (Yale University Library, 2013). Also, many do not have the financial resources, server space, or ability to perform continuous file maintenance and migration activities that are components of digital preservation programs.

Given the current landscape, trade-offs between ease of access and long-term preservation are necessary. In many cases, data dependences on hardware, software, or publisher’s and distributor’s servers create barriers. Information in the digital world is immaterial and must be created each time it is used (Smith, 2004). Obstacles surrounding the ability to provide both ease of access and ensure long-term preservation are linked to data dependencies on hardware, software, or publisher’s (and distributor’s) servers (Smith, 2004, Kirchhoff, 2011). This means that even if a library owns an e-book, it is often impossible to migrate those bytes from one platform to another. In order to facilitate both information accessibility and preservation, “the intellectual content of the book must be unpacked from its reliance on a particular hardware and software and then that content must be securely stowed away and maintained by one or more preservation agencies” (Kirchhoff, 2011, p. 34). Due to the complex relationship between access and preservation, there is currently no e-book solution that “simultaneously meets both the ‘current use’ and ‘future use’ requirements” (Yale University Library, 2013, p. 7). In some cases, it may make economic sense for libraries to purchase an electronic format without thinking about long-term access. Technical manuals, for instance, are in strong demand when current but quickly become obsolete. On the other hand, academic libraries will often purchase titles (e.g., new foreign literature) not because of current user demand but to preserve the content in anticipation of future interest. (Yale University Library, 2013).

In the research community, both ease of access and preservation are vital to the continued development of disciplines, as many “subjects of inquiry and methodologies rely heavily on retrospective [and] current resources” (Smith, 2004, para. 1). For the many stakeholders involved in preservation, access to information removes knowledge barriers and promotes the widespread sharing and vetting of ideas.

### **Preservation Stakeholders**

Today, e-book collections reside in the cloud, meaning that content is stored on server networks that run applications and provide online access to digital resources (GCF Global, 2015). Access to content is gained by purchasing a license, such as a subscription. In the current e-book landscape, there is no guarantee that content will be available long-term – or even next week for that matter. Events like mergers between service providers, evolving licensing models, and the obsolescence of hardware and software can result in the discontinuation of access to countless e-book titles. The instability of e-book collections and packages is a significant barrier as preservation relies on perpetual access to content (and some suggest perpetual functionality as well).

The challenges related to e-book preservation involve a range of stakeholders with very diverse interests, needs, and limitations. They include libraries (both public and academic), publishers, distributors, third-party storage services, and authors. The development of stable and sustainable e-book preservation programs will involve a deep understanding of the perspectives, responsibilities, and roles that each stakeholder brings to the table.

In the library profession, preservation challenges are felt in different ways by public and academic libraries. On the public side, many librarians are tasked with developing collections that meet the present information needs of diverse user communities. Much of this work involves licensing fiction and non-fiction titles, often through services like OverDrive or 3M, released by mass-market publishers who have commercial interests. Traditionally, facilitating long-term preservation has not been a necessary role for public libraries. In many cases, they do not have the legal rights, funds, or data infrastructures required to preserve this content (Ciabattari, 2013) (Kelley, 2012). Some public libraries have begun to challenge this model by negotiating deals with small and non-traditional publishers that allow for e-book files to be mounted on local library servers. While their primary impetus is to gain more control over the content rather than preservation per se, in the current e-book landscape, these libraries may be in the best position to preserve e-books as well.

When discussing the relationship between libraries and for-profit publishers, it is important to keep in mind that “the time horizons of the preservation community and of the commercial sectors are radically different” (Smith, 2004). In the private sector, data retention is often limited to anywhere between five and ten years (Smith, 2004). The time horizon of the preservation community “must include many generations of inquiring humans, not just the next two generations of hardware or software upgrades” (Smith, 2004, para. 35). Some have cited this fundamental difference as the reason why for-profit companies have not played a strong role in the development of preservation systems, even though they provide leadership in areas like information technologies and digital asset management (Smith, 2004). The fundamental difference in data retention schedules is an area where creativity and experimentation will be required to develop preservation programs that promote buy-in from all sides and provide value to all stakeholders involved.

In academia, the ability to guarantee perpetual access to scholarly content is a pressing issue. In the research community, both ease of access and preservation are vital to the continued development of disciplines because inquiry and discovery rely on the utilization of current and retrospective resources (Smith, 2004). In academia, one way that librarians promote preservation is by “brokering deals with publishers and third-party storage services like Portico,<sup>2</sup> which hold scholarly literature in a kind of escrow” (Robertson, 2014, para. 6). Librarians provide funding to such ventures through access fees and are guaranteed access to content when “trigger” events occur – most notably when titles go out of circulation or publishers cease operations (Robertson, 2014).

Due to the complexities and costs of digital preservation repositories, it is unlikely that individual institutions will serve as preservation centers (Smith, 2004). In the current networked environment, “one does not need access to a physical object to have access to information, [and] the relationship between ownership (and physical custody) of information and access to it will be transformed” (Smith, 2004, para. 36). It is likely that a small number of actors, such as Portico, will serve the academic community to preserve the scholarly record (Portico, 2015). In the future, third-party storage services will be an essential part of the “public good information economy that research and teaching have traditionally relied upon for core services such as preservation and collection building” (Smith, 2004, para. 36).

While preservation discussions between libraries, publishers, and third-party storage services will continue for some time, there is a growing volume of titles that are springing up outside of traditional publishing frameworks – they fall into the category of self-published e-books. Last year alone, editors at The Washington Post Book World received 150 self-published books *each day* (Charles, 2014). Established authors like Stephen King, along with those embarking on a writing career, are beginning to turn to services offered through Amazon and Kickstarter to fund or promote their work. This is not a trend that will disappear in the near future. In some cases, corporate publishers are beginning to step into the self-publishing arena as well. In 2012, Simon & Schuster partnered with the self-publishing company Author Solutions, Inc. to form Archway Publishing, a service that assists authors to reach audiences, achieve publishing goals, and make the leap to traditional publishing (Author Solutions, 2012) (Ciabattari, 2013). The preservation challenge surrounding this type of content (besides determining which titles merit preservation efforts) lies in the fact that e-books are currently exempt from mandatory legal deposit. Further, self-published e-books do not require an International Standard Book Number (ISBN), and there is not a simple way to contact or negotiate with independent authors regarding preservation issues (Kelley, 2014). This is a particularly harrowing issue when self-published works have scholarly merit. Currently, there exists a large opportunity for preservationists and computer scientists to work with scholars to develop and document standards for resources that are self-published for research and teaching purposes (Smith, 2004).

## **Market Segments**

If e-books are to be preserved for long-term access, the impetus for action is most likely to come from those institutions that purchase – or license – them. An understanding of market sectors, and in particular of the ways different categories of e-books are acquired by libraries, is thus crucial to understanding the issues, barriers, and potential solutions for e-book preservation.

---

<sup>2</sup> See <http://www.portico.org/digital-preservation/>.



Broadly speaking, the library market for e-books mirrors that for print, with two major sectors: trade books aimed at the general reading public are primarily marketed to public libraries; scholarly works are aimed at the college and university population and acquired by academic libraries. While there is some overlap, and occasional “crossover” titles such as Thomas Piketty’s *Capital in the Twenty-First Century*, most publishers tend to focus on one segment or the other. This division has, thus far, resulted in significant differences in business models and in the terms under which e-books are sold – differences that have profound implications for long-term preservation and access.

While the academic e-book market is still evolving, certain core characteristics seem likely to persist. E-books are sold both as individual titles and in large groups or packages, both directly by publishers and through intermediary distributors, or aggregators, who group titles from many publishers in a common platform. Libraries will license access to some collections on a subscription basis, paying for access year by year, and losing access if payment stops. Many titles, however, are licensed for “perpetual access” with a higher one-time payment. Perpetual access licenses very often include provisions for archiving, either by defining terms under which the library may acquire and use the digital files, or by depositing those files with an agency such as CLOCKSS ([www.clockss.org](http://www.clockss.org)), LOCKSS ([www.lockss.org](http://www.lockss.org)), or Portico. All three of these agencies were founded to preserve e-journals, but CLOCKSS and Portico have since expanded to embrace e-books, with CLOCKSS holding nearly 25,000 e-books (CLOCKSS, 2012) and Portico almost half a million (Portico 2015). These arrangements reflect research libraries’ common interest in long-term preservation of the scholarly record.

By contrast, most trade e-books have been available to public libraries for only the past few years, and the markets and business terms are still diverse and changing rapidly. Again, though, certain common features appear to be emerging. Most trade publishers do not sell e-books directly to libraries, but work through a small number of distributors such as OverDrive, 3M, and Baker & Taylor. While some publishers make titles available for purchase (or, more accurately, perpetual access), it is far more common for e-books to be licensed for a limited term of one or two years, or for a limited number of circulations. Even for those titles available for “purchase,” long-term access depends on the distributors, with no provision for external archiving. Again, these models reflect the practical concerns of public libraries, which primarily collect to meet current reading interests, with a high turnover in collections.

In addition to these two broad patterns, there are of course many variations and experiments. Some public libraries, notably the Douglas County (Colo.) Public Library,<sup>3</sup> working primarily with smaller and non-traditional publishers have developed models for purchasing and mounting the e-book files themselves, thus securing at least the potential to preserve these books. Some academic libraries have become publishers of scholarly e-books, either on their own or in partnership with university presses. These and similar experiments offer potential alternative models for preservation that may in time prove more broadly influential.

For a large and growing number of e-books, however, library markets are virtually non-existent. New types of publication and new methods of publishing are emerging that have not, as yet, found their way into the standard channels by which libraries acquire content. Four main categories can serve as illustrations:

---

<sup>3</sup> [douglascountylibraries.org](http://douglascountylibraries.org)

1. Non-profit publishers: Academic libraries have an interest in many works that are not produced by commercial publishers but by non-governmental organizations, government agencies, and research institutes. Many of these works are now published online, and while freely available on the web, they are not sold to libraries. Although not often thought of as e-books, they have formed an important part of research library collections.
2. Open access publishing: Although the scholarly open access movement has thus far focused largely on journal articles, it is starting to become a force in e-book publishing as well. These works are, by design, freely available, but for that very reason don't fit existing library methods of acquisition and preservation.
3. Non-traditional publishers: The relative ease of distribution over the web and sale through online retailers has brought a proliferation of small publishers and digital offerings from non-publishers (such as advertising agencies, art galleries) offering a mere handful of titles, often available only for individual purchase and download to personal e-readers.
4. Self-publishing: The ease of self-publishing on the web has led to an explosion of hundreds of thousands of titles. Most of these works are again only available for individual purchase and download to e-readers, or as part of e-book subscription services.

For many of these publications the role libraries should play in providing access and the importance of long-term preservation remain uncertain. What does seem clear is that the need for action to secure future access will not wait for that clarity to develop.

### **Issues and Barriers that Require Solutions**

Lack of clarity is pervasive when it comes to e-book preservation. Who should bear the responsibility? How can the costs be supported? What are the technical requirements? What legal rights are involved? Those with the strongest interest in outcomes are not at present in a position to be able to take direct action. Yet, a clearer understanding of the issues and barriers may suggest potential solutions.

#### **Responsibility and Cost**

Primary responsibility for preserving print books has long rested with research libraries: university libraries, the largest public libraries such as New York Public Library, and (in the United States) the Library of Congress. Publishers are supportive of preservation in principle, but are performance driven by financial considerations, and likely to preserve their output only so long as potential sales exceed the cost. Public libraries are primarily responsible for serving the current needs and interests of their communities, and are not funded to sustain access to works once demand for them has faded.

Given the opportunity, scholarly publishers and academic libraries have both invested in e-book preservation, primarily through participation in CLOCKSS and Portico. Costs are shared by contributing publishers and member libraries. Though the economics are not transparent, it seems probable that at least some of the publishers' costs are passed on to libraries and other e-book purchasers in the form of higher prices. Though the long-term costs and success of this model remain to be seen, it offers a reasonable prospect that most mainstream scholarly monographs will be preserved.

The situation with regard to trade books is far murkier. Many popular books of the past – best sellers, dime novels, picture books, mysteries, science fiction and other genre fiction – are now preserved in

academic library collections. With few exceptions however, these libraries didn't acquire such works as they were published, but long after, through donations from private collectors and secondary markets such as used book dealers. This deferred acquisition – and subsequent preservation – is made possible by the first-sale doctrine, under which owners of print books have the right to dispose of them as they choose, lending, donating or selling their copies to others. Many more popular books – from cookbooks to carpentry, self-help to devotional literature – are acquired by the Library of Congress through legal deposit (requiring copies of each new print publication to be deposited with the Copyright Office.)

Currently, however, as noted above, the first-sale doctrine does not apply to e-books. Instead, the rights of owners, even those who have purchased “perpetual access” are limited by licenses, and as noted earlier, few licenses for trade books include any provisions for archiving. Unless this situation changes, research libraries will no longer be able to rely on deferred acquisition to collect and preserve these works. While copyright deposit is mandated for electronic works with no print equivalent, this requirement has thus far only been applied to e-journals, and the potential of the program to provide preservation and long-term access is imperfectly understood. If trade e-books are to be preserved, a new understanding of roles, responsibilities and financing will need to develop, in all likelihood one that actively engages publishers and public libraries, as well as research libraries.

E-books are subject to both legal and technical constraints that limit what libraries can do to ensure their preservation.

### Rights and Obligations

While e-books are subject to the provisions of copyright law, as has been noted most e-book purchases are also governed by licenses that specify what actions the purchaser may take. However, in many cases the license a library signs when purchasing an e-book is only part of the picture. Several parties are involved in the chain of creating, producing, and distributing an e-book, and each may retain certain rights. Here, as elsewhere, the situation is somewhat different for scholarly and trade publications.

Authors of scholarly monographs are generally interested in broad dissemination of their work. The contracts they sign with publishers – often university presses – may assign copyright to the publisher or leave it with the author, and may allow the author certain rights of limited distribution (such as deposit in an institutional repository) but rarely place limits on the publisher's right to distribute the work. Library purchases are often made directly from the publisher, providing a relatively straightforward process for negotiating some form of preservation rights. By contrast, both authors and publishers of trade books depend on sales for their livelihood and survival. They are thus deeply concerned about unauthorized and uncontrolled distribution. The markets, pricing, and royalty structures for e-books are volatile, and all parties involved are cautious about giving away too many rights and suspicious of potential threats.

In both spheres, scholarly and trade, rights issues are further complicated because many library e-book purchases are made not directly from the publisher, but through a distributor such as OverDrive, 3M, ebrary, and many more. The library's rights to the content are controlled by its license with the distributor, while the distributor's rights are in turn limited by its contracts with many publishers, whose own rights are determined by agreements with hundreds or thousands of authors. In such an environment it is no wonder that the final link, between library and distributor, is generally quite limiting and cautious.

Complicated as this may seem, e-book licenses for major publishers both trade and academic at least offer opportunities for negotiation with high-volume impact. For the output of small independent publishers, self-published authors, and works on the open web, the situation is much worse. Many small publishers' and authors' works are for sale only through major on-line retailers – Amazon, Apple, and Barnes & Noble. As we've seen, they are marketed to individual consumers, whose rights are governed by those retailers' standard contracts, leaving little-to-no room for library negotiation of preservation, or indeed any other, terms. Any attempts to negotiate directly with authors and publishers will necessarily be scattered and diffuse.

We might then turn with a sigh of relief to a class of e-books that are freely available for download from the web. While many of these are not what we typically think of as e-books – research reports from scientific institutes, policy briefs from think tanks and other non-governmental organizations – they are of considerable interest to academic libraries. Even here, though, rights issues are not absent. Most of these works are subject to copyright and, unless they are made available under some form of Creative Commons license, the ability to download a copy for personal use does not automatically convey the right to make further copies for preservation and access.

### **Technical Requirements**

Finally, securing the right to preserve an e-book does not necessarily include the ability to do so. Publishing formats for e-books are diverse, and attempts to standardize on a single format such as EPUB3 have thus far met with limited success. Many are produced in multiple formats for use on different e-readers; preserving a single format may or may not suffice for future use. Far more than with journal articles, e-books are aesthetic objects, where style and appearance are considered important aspects of the reading experience; preservation of the words and images is necessary, but not always sufficient. Moreover, the e-books that are most at risk – those without print equivalents – are often the most challenging to preserve, incorporating multiple media and experimental features. DRM software is often part of the standard distribution package, requiring a separate format for preservation purposes.

Finally, some e-books, such as the *Interaction of Color*, are beginning to be produced as apps, with the application software an integral part of the work. From a technical standpoint, these objects might better be viewed as a software preservation challenge, rather than as e-books per se. Conceptually and intellectually, however, they seem more related to books than other types of software.

A relatively small number of libraries are likely to have the wherewithal, in technical expertise, storage capacity and back-up storage to preserve and render back e-books from diverse sources on a large scale. Just as with e-journal preservation, concentration of effort and collaboration around agencies such as Portico and CLOCKSS are likely to be the norm.

### **Recommendations for Action**

For years to come, attempts to preserve e-books are likely to prove messy, uncoordinated, inconclusive, and incomplete, yet that is still better than waiting for things to become simpler and clearer. Failed attempts and partial successes can help narrow the range of possible solutions. Small-scale experiments may address specific technical, legal, or financial challenges, and then build towards more systematic solutions. Engaging multiple stakeholders in different ways can help foster a sense of collective

ownership of the issues. The recommendations that follow are offered, not as resolutions, but as explorations.

### Focus on Outcomes

Ultimately, preservation is about ensuring the survival of e-books so that they may continue to be available for future use. However, different types of content may suggest different types of use, by different groups, at different times. For any group of e-books, libraries will need to consider:

- Who is likely to want access in future? For what purposes? When?
- Is preservation of the content (text and images) sufficient? How important are other features?
- Are there classes for which a semi-dark archive, with limited or embargoed access, is sufficient?

Focusing on one standard, ideal model of preservation for all types of e-books is likely to impede incremental progress that may ultimately get us closer to the ideal.

### Build Consensus on Priorities

Numerically and financially, a high percentage of the e-books purchased by libraries are still also produced in print and preserved in that form. At the other extreme, many works produced only as e-books are not purchased by libraries at all. With limited resources to devote to e-book preservation, should libraries focus on value (protecting investment, broad interest) or risk (ephemeral interest, entrepreneurial publishing)? Should they seek quick results (low-hanging fruit), or concentrate on difficult questions with potential for broad impact? Different stakeholders may well answer these questions differently. Finding like-minded partners may help organize action on multiple fronts.

### Clarify Roles

There is broad consensus within the library community, and at least in principle among publishers, that preservation is important, but little discussion of who can and should take responsibility for different aspects – negotiating rights; storing content; maintaining format integrity; migrating access; supporting costs. Here again, different stakeholders are likely to have different perspectives, and roles may vary for different types of e-books. Two types of action are needed:

- Sectors within the library community organized by type (public, academic), size (large, small) and function (preservation, collection development) need to organize discussions and build consensus around willingness to act and effective locus of activity.
- Libraries, individually and through organizational channels, need to propose a variety of potential models to publishers and providers of e-books to test their willingness and ability to support specific preservation actions.

### Experiment

A major barrier to e-book preservation is that it has been difficult to see how anyone can take action without massive changes in the way e-books are licensed, distributed, and stored. While systemic change is still difficult, small scale experimentation is possible.

- As noted earlier, CLOCKSS, LOCKSS, and Portico are already preserving tens of thousands of e-books. Individually and collectively, libraries can take a more active role in encouraging new publishers to participate in these programs.
- New open-access ventures such as Knowledge Unlatched have secured agreement from HathiTrust, Portico, and others to preserve and provide access to their e-book publications, a model that might be extended to additional content (HathiTrust, 2014).
- Some of the e-books most at risk are the output of innovative, start-up ventures (some of which have already suspended activity). These firms might be amenable to a “rescue” operation in which one or more libraries would secure the right to preserve (and possibly provide access to) their e-book content.
- Research libraries might explore a model used for some print archives, in which a particular publisher’s e-book output would be deposited with the library for preservation, with access embargoed for an interval, and/or restricted to researchers on the archive’s premises.
- Many self-published books may ultimately be of more interest as a corpus than for reading as individual titles. Libraries might seek agreement from providers of self-publishing platforms to secure and preserve the content for non-consumptive use.

These examples are cited merely to suggest a range of possibilities. Ultimately, none may prove viable or extensible, but as with e-books themselves, innovation will require a willingness to fail sometimes.

### Reward Success

If innovation requires a willingness to fail, it is also true that limited successes, however imperfect, must be encouraged. Publishers and libraries who demonstrate a willingness to take risks in order to preserve e-books ought to be recognized, with encouragement to others to follow and build on their efforts. Rewards may take many forms – favorable notice in library forums and media; formal awards and citations; and, most tangibly, increased business for those most committed to preservation. This is not to suggest that all initiatives should be received uncritically, but that acknowledgment of good-faith effort should accompany critical analysis, and that the best form of criticism is to point the way to something better.

### Conclusion

At a time when the library profession is in transition due to the proliferation of technology in society, one may ask why time and resources should be invested in a complex challenge that addresses the future state of library collections. Are we borrowing trouble when we should be focused on today? In reality, the policies that guide e-book acquisition decisions can be coupled with preservation strategies. The digital universe is slowly erasing the distinction between collecting content “just in case” (i.e. an item proves later to be valuable) and “just in time” (i.e. for use today). As technologies evolve, digital items need to be optimized for preservation at the time of their creation, the time of their deposit into repositories, and

managed over time (Smith, 2004) to guarantee long-term access in an environment where hardware and software obsolescence is guaranteed.

The good news is that we are in the early days of e-book collection development. There is time to develop strategies and frameworks that support the preservation of e-book content and functionality. At the same time, the knowledge that digital collections are not permanent signal that we must take action and begin experimentation. In the future, there will be “no uniqueness, no scarcity, [and] no category of ‘rare’” (Smith, 2004, para. 45) when dealing with digital information. These resources are part of our national cultural heritage, and the responsibility for ensuring their survival must be shared. Publishers, foundations, and the federal government must all play a part. But, action will only begin if librarians take a more active role in the stewardship of collections and build effective relationships with external stakeholders to collectively evaluate information and determine what items best represent the cultural heritage of the early 21<sup>st</sup> century. The answers will not be simple, but as evidenced through advances in e-journal preservation, they are not impossible.

As librarians at Columbia University Libraries grappled with the preservation questions surrounding the *Interaction of Color* app, we had an opportunity to discuss preservation challenges with Yale University Press. We learned that publishers are also grappling with these issues, and there are opportunities to work together, experiment, and discuss frameworks related to the preservation of enhanced e-books. For instance, one strategy may be to begin with the preservation of underlying images and text. While this excludes functionality, it provides a starting point that may better acquaint information professionals and publishers with the challenges at hand. As we learn from each other, we move towards sustainable, long-term preservation solutions one byte at a time.

## References

- ARL. (2005, October). *Urgent Action Needed to Preserve Scholarly Electronic Journals*. Retrieved from [http://www.arl.org/storage/documents/publications/e\\_journal-preservation-15oct05.pdf](http://www.arl.org/storage/documents/publications/e_journal-preservation-15oct05.pdf)
- Armstrong, C., & Lonsdale, R. (2009, November). *E-book collection management in UK university libraries: Focus groups report*. Information Automation Limited, Final Report. Retrieved from <http://observatory.jiscebooks.org/reports/e-book-collection-management-in-uk-university-libraries-focus-groups-report/>
- Author Solutions. (2012, November 27). Simon & Schuster and Author Solutions launch Archway Publishing. Retrieved from <http://www.authorsolutions.com/News-Events/News-Releases/2012/Simon-and-Schuster-and-Authors-Solutions-Launch-Archway-Publishing/>
- Billington, J. (2013). Federal agency and federal library reports. In D. Bogart (Ed.), *Library and book trade almanac* (pp. 55 - 77). Medford: Information Today, Inc.
- Blummer, B., & Kenton, J. (2012). Best practices for integrating e-books in academic libraries: A literature review from 2005 to present. *Collection Management*, 37(2), 65-97.
- Center for Research Libraries. TRAC and TDR Checklists. Retrieved from <http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0>

- Charles, R. (2014, October 1). No, I don't want to read your self-published book. *The Washington Post*. Retrieved from <http://www.washingtonpost.com/blogs/style-blog/wp/2014/10/01/no-i-dont-want-to-read-your-self-published-book/>
- Ciabattari, J. (2013, September 3). Now there are 5. *Library Journal*. Retrieved from <http://lj.libraryjournal.com/2013/09/publishing/now-there-are-5/#>
- CLOCKSS. (2012). *Home*. Retrieved from <https://www.clockss.org/clockss/Home>
- GCF Global. (2015). *What is the cloud?*. Retrieved from <http://www.gcflearnfree.org/what-is-the-cloud>
- Gerden, E. (2014, January 9). Russian e-book market matures, shows more potential for growth. *Publishing Perspectives*. Retrieved from <http://publishingperspectives.com/2014/01/russia-is-now-the-worlds-third-largest-ebook-market/>
- HathiTrust. (2014, April). HathiTrust rights and access working group charge [Web log message]. Retrieved from [http://www.hathitrust.org/rights\\_and\\_access\\_charge](http://www.hathitrust.org/rights_and_access_charge)
- Kelbanoff, A. (2013, December 16). The tantalizing English language eBook opportunity in China. *Digital Book World*. Retrieved from <http://www.digitalbookworld.com/2013/the-tantalizing-english-language-ebook-opportunity-in-china/>
- Kelley, M. (2012, October 18). Random House says libraries own their ebooks. *Library Journal*. Retrieved from <http://lj.libraryjournal.com/2012/10/opinion/random-house-says-libraries-own-their-ebooks-lj-insider/>
- Kelley, Michael. (2014, October 3). How libraries preserve e-books. *Publishers Weekly*. Retrieved from <http://www.publishersweekly.com/pw/by-topic/industry-news/libraries/article/64271-check-it-out-with-michael-kelley-how-libraries-preserve-e-books.html>
- Kirchhoff, A. (2011) E-books: The preservation challenge. *Against the Grain*, 23(4). Retrieved from <http://docs.lib.purdue.edu/atg/vol23/iss4/10/>
- Kirchhoff, A. & Morrissey, S (2014, June 1). *Technology Watch Report 14-01: Preserving e-books*. Retrieved from <http://www.dpconline.org/advice/technology-watch-reports>
- Portico. (2015). *Our Organization*. Retrieved from <http://www.portico.org/digital-preservation/about-us/our-organization>
- Robertson, A. (2014, May 9). The fight to save endangered books. *The Verge*. Retrieved from <http://www.theverge.com/2014/5/9/5688146/the-fight-to-save-endangered-ebooks>
- Rüdiger Wischenbart Content & Consulting. (Spring 2014). *Global eBook: A Report on Market Trends and Developments*. Retrieved from [www.wischenbart.com/upload/123400000358\\_04042014\\_final.pdf](http://www.wischenbart.com/upload/123400000358_04042014_final.pdf)
- Smith, A. (2004). Preservation. In S. Schreibman, R. Siemens & J. Unsworth (Eds.), *A Companion to Digital Humanities*. Oxford: Blackwell.



- Utsumi, I. (2014, March 10). *The Brazilian book market*. The Brazil Business. Retrieved from <http://thebrazilbusiness.com/article/the-brazilian-book-market>
- Vyas, H. (2014, September 11). India's ebook industry shows great potential. *Digital Book World*. Retrieved from <http://www.digitalbookworld.com/2014/indias-ebook-industry-shows-great-potential/>
- YaleNews. (2014, May 14). 'Interaction of Color' app wins George Wittenborn Memorial Book Award [Web log message]. Retrieved from <http://news.yale.edu/interaction-color-app-wins-george-wittenborn-memorial-book-award>
- Yale University Library. (2013, March 3). *The eBook Strategic Plan Task Force*. Retrieved from <http://www.library.yale.edu/departments/collection-development/Yale-ebook-task-force-rpt.pdf>