Copyright, Digitization, & RFID

Follow this and additional works at: https://scholarsarchive.library.albany.edu/jlams

Part of the Library and Information Science Commons

Recommended Citation

Available at: https://scholarsarchive.library.albany.edu/jlams/vol8/iss1/1

This Article is brought to you for free and open access by Scholars Archive. It has been accepted for inclusion in JLAMS by an authorized editor of Scholars Archive. For more information, please contact scholarsarchive@albany.edu.
Library and Administration Management Section (LAMS) Officers

President
Julie Johnson

President Elect
Tim Burke

Past President
Rebekkah Smith Aldrich

Secretary
Patricia Kaufman

Treasurer
Mary M. De Bellis

JLAMS
Richard J. Naylor

Web Liaison
Frank Rees

At Large Board Members
Rebecca Lubin
Ed Falcone
Gillian Thorpe
Claudia Depkin
Rob Carle

To submit an article read the JLAMS editorial policy on our website at:
Then send an eMail proposal to:
naylor@ColonieLibrary.org

New York Library Association
6021 State Farm Road
Guilderland, NY 10284
518-432-6952/800-252-NYLA email: info@nyla.org web: www.nyla.org

President
Marcia Eggleston
315-250-0352
megglest@nncsk12.org

Executive Director
Michael J. Borges director@nyla.org

CONTENTS

3 President’s and Editor’s Message

4 How Copyright Theory Affects Practices:
a Primer For Information Professionals
by Robert Weiss and Katherine M. Shelfer.

25 RFID Materials Circulation and Handling:
A Model for Improving Customer Service
By Christine McDonald

31 Digitizing Everything? Part II: Piloting
Metadata Creation By Kimmy Szeto

https://scholarsarchive.library.albany.edu/jlams/vol8/iss1/1
JLAMS, the electronic Journal of the Library Administration and Management Section of the New York Library Association, begins its eighth year, and we are privileged to introduce the Fall 2011-12 JLAMS.

JLAMS provides a valuable outlet for the dissemination of articles, academic papers, and essays of interest to administrators and managers of all types of libraries: academic, public, school and special libraries. As administrators and managers, we have a lot in common, but we have few places to share what we know. JLAMS was the first peer-reviewed journal in NYLA, and the goal was to set a high standard for future publications. Readers of JLAMS are well-served by our team of referees, as are those whose contributions are published here. Submissions are vitally important. For information on article submissions, editorial policy, a submission form and more, visit the JLAMS website page at http://www.nyla.org/page/jlams-journal-of-the-library-administration-and-management-section-224.html.

Over the seven plus years that we have been publishing JLAMS we have enjoyed working with many interesting colleagues. This month we have three articles that run the gamut from a state of the art digitization project at Queens, to an experiential article on RFID. Our lead article is an in depth look at copyright which in light of our issues with e-book publishers seems very timely. We hope you enjoy the articles and that you consider writing one yourself and sharing new information with colleagues for the betterment of all our libraries.

JLAMS is made possible by NYLA membership. LAMS receives funding based upon the number of people who select LAMS as their primary NYLA section, as well as by those who pay an additional $7.00 to add LAMS as a secondary section. Please keep this in mind when renewing your NYLA membership. And thanks for your support!
HOW COPYRIGHT THEORY AFFECTS PRACTICES:

A PRIMER FOR INFORMATION PROFESSIONALS

By Robert Weiss and Katherine M. Sheller.

Abstract: Over the course of several decades the U.S. economy has evolved from a manufacturing base to an information base. Concomitant with this transition, intellectual property theories — which define the ownership and appropriate usage of information — have become increasingly important to assess, and IP practices are now mission critical to predict and understand. Copyright law reflects ambiguities and conflicts. This paper identifies and clarifies theory in order to help librarians and other information workers understand and manage copy rights (in the generic sense) and copy privileges. It first examines copyright theory and its historical and contemporary impact on copyright law and information-sharing practices. It then explores historical changes in the legal and social frameworks of copyrights, with a special focus on policies and practices that control and manage these rights. It concludes by identifying emerging trends, policy considerations, and best professional practices that relate to the library profession.

SECTION ONE: COPYRIGHT FUNDAMENTALS

The digital revolution is changing society’s fundamental definitions of property as well as its attitudes toward copy rights and copy privileges. Librarian Rebecca Butler accurately describes copyright as ‘a very muddy area of law with lots of interpretations’ (2004, p. 3). The ambiguities and conflicts inherent in copyright law underscore the need to identify and clarify fundamental concepts associated with copyright. Our research suggests many reasons for copyright infringement—e.g., infringers (a) assume...
they won’t get caught, (b) ignore or dismiss ethical/legal ramifications, (c) take calculated risks for financial and other rewards, and/or (d) redefine ‘fair use’ to favor their use.

Although the U.S. Copyright Act excepts innocent infringement of protected copyrights, the following example demonstrates that current bundles of copyrights and copy privileges involve an increasingly complicated set of opportunities, practices, expectations, and economic and judicial impacts.

**Example 1.** In 2007, Stephanie Lenz uploaded to YouTube a video of her young children dancing to Let’s Go Crazy by Prince. Universal Music directed YouTube to remove the video, claiming it was an infringement of Universal’s copyright. When YouTube complied, Lenz challenged Universal, arguing that her videotape constituted a legitimate or fair use of the music under existing copyright law. When Universal disputed this claim, Lenz sued the company (Lenz v. Universal Music Corp.).

In this section, we introduce and discuss four issues in a contemporary context: (1) existing copyright laws; (2) private versus public use; (3) fair use; and (4) digital rights management.

**1. Existing Copyright Laws**

Title 17 of the U.S. Code defines a copyright as ‘a type of protection granted by federal law to the authors of literary, musical, artistic, dramatic, and other types of intellectual works.’ Section 101 states that copyrights are exclusive but are also ‘limited by statute to a set number of years, after which the work becomes part of the public domain and may be used by anyone.’ Section 106 lists the six exclusive rights of copyright owners:

1. ‘To reproduce the copyrighted work in copies or phonorecords;
2. To prepare derivative works based upon the copyrighted work;
3. To distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending;
4. In the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works, to perform the copyrighted work publicly;
5. In the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including the individual images of a motion picture or other audiovisual work, to display the copyrighted work publicly; and
6. In the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission.’

These rights have been the source of extensive controversy and conflict that can be traced as far back as the invention and use of the printing press (refer to Section Two). The current embodiments of these rights (in U.S. law) suggest that certain aspects of the official definition (cited above) merit consideration. Example 2 (below) frames the discussion of precepts that follows.

**Example 2.** John Doe applies for a copyright to protect his ownership of his thesis on the role of information literacy within the field of librarianship. Although he may be awarded a copyright for his thesis, he will not be granted a copyright for either the concept of information literacy or the concept of librarianship. These concepts are considered ideas and therefore cannot be copyrighted by anyone.
Copyright applies exclusively to intellectual works.

Ideas are the nonphysical building blocks of human progress—e.g., they trigger private thought, change society’s accepted behaviors, and redirect institutionalized resources/responses. Ideas are not subject to the limitations associated with physical property, so the term intellectual property is generally used to clarify this distinction.

Attaching ownership to ideas is problematic.

U.S. copyright law explicitly stipulates that ‘copyright protection extends only to the particular form or expression in which an author conveys ideas, information, etc.’ but not to the ideas themselves (U.S. Copyright Act, Section 101). This stipulation involves a fundamental distinction between theory and practices.

Theory

It is possible to distinguish between an idea and the expression of that idea. For example, the German philosopher Johann Gottlieb Ficht differentiated Inhalt, the actual content, from Form, the writer’s individualized expression. Ficht posited that the former should be available to everyone, whereas the latter should remain the writer’s personal possession (Kawohl & Kretschmer, 2009).

Practice

The boundaries between theory and current practices are amorphous. Protections are quite strong, but somewhat idiosyncratic—e.g., ‘sweat of the brow’ replication is generally excluded, yet reverse-engineered adaptations and publications known as ‘formula’ fiction are generally eligible for copyright protection. In the context of Example 2, given the plethora of published work on John Doe’s thesis topic, established criteria are used to determine if his work is sufficiently distinctive to be awarded a copyright.

The official definition of copyright explicitly asserts that copyright protection is granted by federal law.

The U.S. Constitution empowers Congress to award ‘for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries’ (Article I, Section 8). This wording challenges the longstanding belief, prominent in Europe, that copyright is a ‘natural’ right—i.e., that copyright is an inherent feature of the creative process.

The explicit purpose of both copyrights and patents is to encourage the dissemination of ideas and technology.

It is important to note that the constitutional authorization of copyright contains no references to protecting property or providing anyone with a source of income. According to Anthony Falzone of the Stanford Law Center for Internet and Society (CIS), ‘The whole point wasn’t to protect stuff. It was to encourage people to make stuff’ (quoted in
Liptak, 2011). This fact explains why the Congress explicitly restricted copyrights to a defined length of time. However, since the enactment of the first U.S. copyright law in 1790, Congress has steadily and dramatically expanded copyright duration (see Table 1 below). Opponents have (mostly unsuccessfully) challenged these extensions on the grounds that they limit public access to information.

- **The framers of the Constitution believed that denying authors and inventors any rights to their works would discourage creativity.**

The framers granted creative individuals a temporary — and initially very short-term — monopoly designed to promote the public good. As copyright scholar Siva Vaidhyanathan explains, ‘Incentive, not property or natural law, is the foundational justification for American copyright’ (2001, p. 45). For this reason, he suggests that the term *copyprivileges* is more expressive of the intentions of the system envisioned by the Founding Fathers (p. 21).

Unfortunately, over the course of more than two centuries, the constitutional basis for awarding copyrights and patents has largely disappeared from the public debate. This development has generated continual conflicts between public and private uses and what legally constitutes ‘fair use’ of copyrighted materials.

2. Public and Private Use

As illustrated in Example 1 (above), Stephanie Lenz’s private use morphed into a public performance of another party’s copyrighted song. This case reflects the highly troublesome issue of differentiating public use from private use.

- **Copyright restrictions are ‘not unlimited.’**

Copyright laws and policies intended to balance the conflicting rights of (a) creators/owners and (b) consumers/users can be traced back to the early Middle Ages. At one extreme of this ongoing conflict are ‘the many content producers [who] consider they have an absolute right to protect all access to and uses of their work’ (Samuelson, 1998, p. C3). At the other extreme are consumers who believe they are entitled to unrestricted access to and use of any resource they can obtain. Example 3 below provides evidence of the widespread acceptance of this perspective.

**Example 3.** A recent congressional report indicated that ‘80% of consumers admit they regularly buy fake or pirated products, with little remorse or concern about the consequences of those purchases’ (cited in WIPO, ‘Counterfeiting’).

- **Copyright simultaneously performs both a restrictive role and a protective role.**

Restrictions promote and reward creativity. Protections are essential to sustain a free, open society. This underlying tension constitutes the most contentious aspect of copyright. Shelfer observes that ‘the free transmission of ideas is considered to be critical for human progress as well as for the perpetuation of an open society.’ For this reason, she concludes that ‘copyright and the First Amendment are mutually incompatible’ (1994, p. 330).
3. Fair Use

Copyright legislation includes provisions for fair use, which means ‘using …copyrighted material in a reasonable manner without [the copyright owner’s] consent, notwithstanding the monopoly granted to the owner’ (Posch, 1990). The fair use guidelines that evolved over time were eventually codified in the Copyright Act of 1976. This law generally applies to activities such as ‘criticism, comment, news reporting, teaching, scholarship, and research’ (U.S. Copyright Office, Fair Use). Section 107 identifies four factors that must be considered in any official decision concerning fair use:

1. ‘The purpose and character of the use, including whether such use is of commercial nature or is for nonprofit educational purposes [the test of spontaneity]

2. The nature of the copyrighted work

3. The amount and substantiality of the portion used in relation to the copyrighted work as a whole [the test of brevity]

4. The effect of the use upon the potential market for, or value of, the copyrighted work’ [the test of cumulative effect].

Significantly — and in contrast to what some users might believe — fair use does not consider either cost avoidance or convenience. For example, it is not normally considered fair use to (1) create, replace, or substitute for anthologies, compilations, or collective works; (2) avoid the cost of purchasing ‘consumables’—e.g., workbooks, exercises, standardized tests, and test booklets and answer sheets; (3) avoid purchase of books, publishers’ reprints, or periodicals (see, e.g., American Geophysical Union v. Texaco Inc., 1994); or (4) create an ongoing enterprise that results in a net positive income for providing unauthorized access to copyrighted works (meaning to charge a fee larger than the actual cost of the copy). This practice was the Achilles heel of campus copy centers and similar establishments (see, e.g., Books Incorporated v. Kinko’s, 1991).

4. Digital Rights Management

For most of human history, the expression of ideas has taken a discrete (monolithic) physical form—e.g., manuscript, printed book, DVD. Today, however, ideas are increasingly expressed and transmitted in dynamic digital formats. As a result, ‘Users find that they can edit video on their computers and can take movies from their DVDs and “mash them up” with other media and upload them to YouTube … [so] fundamental distinctions between users and creators are rapidly becoming obsolete’ (Postigo, 2008, p. 1014). Some mashups are predictable (see Example 1 on Stephanie Lenz, above). However, copyright is not—and never was—driven by the discrete forms and formats associated with monolithic expressions of an idea. Rather, it is best conceived of as ‘a “bundle” of rights that includes …the exclusive right to make copies, authorize others to make copies, create derivative works such as translations and displays in other media, sell the work, perform the work publicly, and petition a court for relief in case others infringe on any of these rights’ (Vaidhyanathan, pp. 20–21).

Critics of copyright laws contend that intellectual property differs from tangible property in fundamental ways.

- **Intellectual property is nonrivalrous.** Multiple parties can use IP at the same time (Clement, 2003). In contrast to physical property, then, ‘information can be transferred without leaving the possession of the original owner’ (Barlow, 1994, pp. 7–8).
Reproducing and utilizing intellectual property does not diminish its value. In fact, the converse might be true: The more widely an idea is disseminated, analyzed, and developed, the greater the chances that it will be used in a socially beneficial manner.

Intellectual property is not physical. Because IP is nonexcludable, it is extremely difficult to contain (Menell, 2007, p. 726). According to one prominent fair use proponent, ‘Information that isn’t moving ceases to exist as anything but potential’ (Barlow, 1994, p. 7).

SECTION TWO: EXPANSION OF COPYRIGHT PROTECTION IN AMERICA

How did copyright law shift from ‘public good’ to ‘private property’? Our research suggests that the republican philosophy of Colonial and Early America shifted over the course of the nineteenth and twentieth centuries because copyright owners successfully promoted the right of creators to control their creations. This right became conflated with traditional property rights in the concept we currently label intellectual property. The following discussion is supported by Table 1 and Table 2, below:

Table 1. Extensions of Copyright Protection

<table>
<thead>
<tr>
<th>LAW</th>
<th>COPYRIGHT DURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copyright Act of 1790</td>
<td>14 years</td>
</tr>
<tr>
<td>Copyright Act of 1909</td>
<td>28 years, renewable for 28 more years</td>
</tr>
<tr>
<td>Copyright Act of 1976</td>
<td>Author’s lifetime plus 50 years</td>
</tr>
<tr>
<td>Copyright Term Extension Act of 1998</td>
<td>Author’s lifetime plus 70 years</td>
</tr>
</tbody>
</table>

Table 2. Summary of U.S. Copyright Laws and Legislation

<table>
<thead>
<tr>
<th>YEAR</th>
<th>DEVELOPMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1865</td>
<td>Congress extends copyright protection to photographic prints and negatives</td>
</tr>
<tr>
<td>1870</td>
<td>Congress applies copyright to foreign translations of U.S. literary works</td>
</tr>
<tr>
<td>1909</td>
<td>Congress establishes ‘a copyright against the unauthorized mechanical reproduction of musical compositions,’ applied primarily to phonographic records and pianola rolls (Goldstein, 1994, p. 67)</td>
</tr>
<tr>
<td>1912</td>
<td>Congress extends copyright protection to movies</td>
</tr>
</tbody>
</table>
Eighteenth Century

In colonial America, the right to publish was central to the concept of the public sphere, defined by historian Michael Warner as an arena in which ‘political discourse could be separated both from the state and from civil society’ (1990, p. x). According to Warner, such discourse generally reflected republican sentiments. Colonial American writers and philosophers argued that despotism and tyranny survive by restricting the flow of ideas in order to keep people ignorant. Thus, the democratization of print is inextricably linked to the democratization of power (p. ix). As Thomas Jefferson explained, ‘That ideas should freely spread from one [individual] to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature’ (quoted in Barlow, 1994, p. 1).

When the United States achieved independence, it defined free speech as a fundamental right. It is not coincidental that the First Amendment to the Constitution prohibits Congress from ‘abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.’ Despite these convictions, however, most of the nation’s leaders conceded that some form of copyright was necessary. Consequently, as previously mentioned, the drafters of the Constitution incorporated a clause that empowered Congress to grant copyrights and patents in order to encourage the dissemination of ideas and inventions.

In May 1790, Congress passed the first U.S. Copyright Act. Reflecting the Statute of Anne (a landmark English copyright law passed in 1710), the Copyright Act granted authors control of their products for 14 years. This law applied primarily to books, but it also protected maps and charts. Congress has consistently revisited its copyright policies, with mixed results, to revise and extend copyright protection to all new forms of media.

Nineteenth Century

During the course of the nineteenth century, authors and content owners began to intertwine the concept of copyright with the concept of private property, a core social value. Proponents of this philosophy defined the expressions of ideas and creativity as intellectual property, arguing that authors own their work(s) in the same ways that people own tangible properties—e.g., their homes and possessions. It follows logically, then, that intellectual property should be subject to many of the same rules and restrictions as physical property. Going further, the unauthorized use of intellectual property constitutes a form of theft. The ascendancy of the intellectual property philosophy is evidenced by the gradual extension of the duration of copyright protection, as illustrated in Table 1 (above).

As the duration of copyright protection increased, so did its scope. Beginning in 1865, Congress has consistently awarded copyrights to nonprint media. Significantly, the Supreme Court has formally upheld this policy. In Burrow-Giles Lithographic Co. v. Sarony, summarized in Example 5 (below), the Court expanded the definition of ‘author’ to encompass ‘he to whom anything owes its origin; originator; maker; one who completes a work of science or literature.’

Example 5. Photographer Napoleon Sarony sued the Burrow-Giles Lithographic Company for copyright violation because it had reproduced his photograph of Oscar Wilde without requesting his permission. The Court ruled in favor of Sarony, asserting that a photograph constituted an original work of art (Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 1884).
Another major development occurred in 1870, when Congress amended the Copyright Law to transfer copyright registration authority from the federal district courts to the Library of Congress. From this point on, copyright became an essential feature of the library environment in the United States.

**Twentieth Century**

Over the course of the twentieth century, intellectual property arguments gradually displaced the republican ideals reflected in the Constitution. One reason for this transformation was the emergence of large media enterprises such as the Recording Industry Association of America (RIAA) and the Motion Picture Association of America (MPAA). Individual artists and authors increasingly assigned their copyrights to these establishments, in exchange for what became substantial investments in publishing and promoting their works in a wide variety of formats and venues (see, e.g., Fisher, 2004, Chapter 2).

As a result, the conglomerates that support artists and authors now have a vested economic interest in protecting and extending copyrights. (Recall that in Example 1, Universal Music, not Prince, accused Stephanie Lenz of copyright infringement.) These developments might explain why the past few decades have witnessed an onslaught of legal actions, legislation, international treaties, and public debates, accompanied by a vast literature generated by technology writers, librarians, legal scholars, and other concerned parties.

Prior to 1980, the laws and judicial rulings of the previous 200 years, however contentious, nevertheless had established a balance between authors’ rights and public access. However, the emergence of digital information, computer networks, and the Web, coupled with the rapidly expanding integration of modern communication and information technologies into people’s everyday lives, drastically upset that balance (see, e.g., National Resource Council, 2000). Information scholar and lawyer Anne W. Branscomb identified three factors that were primarily responsible for the breakdown of the traditional system of copyright protection:

1. ‘The development of new information and communication technologies
2. The globalization of the marketplace

**The Digital Age and the Twenty-first Century**

The so-called ‘digital age’ can be traced back to about 1980. In that year Congress amended the Copyright Act to include software (Branscomb, 41). This amendment received judicial approbation in 1986 in *Whelan Associates, Inc. v. Jaslow Dental Laboratory, Inc.* (797 F.2d 1222). In this case a federal appeals court ruled that the structure or sequence of a software program constituted an expression of an idea rather than the idea itself. Consequently, it was protected by the Copyright Act. Since that time, an almost breathtaking succession of communication and information technologies and the Internet have extended into every facet of people’s lives. ‘The digital revolution,’ argues information scholar Peter Menell, ‘has displaced General Motors and other manufacturing enterprises from the top of the economic food chain’ (2007, p. 716).

In the course of this development, information has morphed from a product into a process. This transformation rendered traditional copyright policies both unfeasible and unenforceable (Branscomb, 1988; Barlow, 1994). As Branscomb observed, ‘The attempt to force these new technologies into outmoded categories can create absurd and contradictory situations that threaten to undermine public confidence in the principle of intellectual property rights itself’ (p. 40). In the absence of a viable
overriding philosophy of copyright protection, conflicts have erupted every time a new technology is introduced to the market. These disputes inevitably found their way into the political arena. Table 3 summarizes some of the key judicial rulings that emerged from these conflicts.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Laws and Rulings</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home videorecordings</td>
<td><em>Sony v. Universal</em> (1984)</td>
<td>Using a videorecorder (e.g., Sony Betamax) to make private recordings for time shifts (e.g., a program that airs when you are at work) constitutes fair use and therefore does not qualify as a copyright infringement (Goldstein, 1994, pp. 144–159).</td>
</tr>
<tr>
<td>Databases</td>
<td><em>Feist Publications v. Rural Telephone Service Co.</em> (1991)</td>
<td>To qualify for copyright status, a database must utilize the data in a creative manner. This ruling ‘sounded the death knell for the sweat of the brow doctrine’ (U.S. Copyright Office, 1997).</td>
</tr>
<tr>
<td>Reverse engineering</td>
<td><em>Sega v. Accolade</em> (1992)</td>
<td>Reverse engineering — the process in which users work backwards to ‘unlock’ the product’s technology in order to modify it — constitutes fair use when the user has a legitimate interest in accessing the functional elements of a technology; for example, to create a similar product. Significantly, the court cited the <em>Sony</em> decision as a precedent (Goldstein, 1994; Sheller, 1994.)</td>
</tr>
<tr>
<td>Computer interfaces</td>
<td><em>Apple v. Microsoft</em> (1992)</td>
<td>Apple could copyright its specific graphical user interface (GUI), but it could not copyright the concept of a GUI (i.e., its essential features). Further, the ‘similarities’ in the functional elements in various GUIs — particularly Microsoft Windows — represent ‘standardization,’ not ‘unlawful copying’ (quoted in Goldstein, 1994, p. 208).</td>
</tr>
</tbody>
</table>

**Table 3. Landmark Copyright Cases**

Perhaps the most transformative technology to emerge from the digital age is the Internet. Initially established to promote the free, unimpeded flow of information, the Internet rapidly evolved into a highly profitable economic enterprise (Fayenson, 1999, p. 9). As large stores of data became available on the Internet and the World Wide Web (WWW), cyberspace inevitably became a major arena for intellectual property wars.

**Domestic (U.S.) Piracy**

Today, simple technologies facilitate the widespread illegal reproduction and distribution of software, music, videos, and other media. Media owners came to refer to these practices as piracy, a term once reserved for the crimes perpetrated by sea villains. Significantly, in addition to expanding access to the
products of modern technology, piracy also enriches organized criminal enterprises. To combat piracy, media owners and technology creators implemented two risk-management strategies:

1. **Content owners use litigation as an economic disincentive.** Record companies have sued thousands of individuals for illegally downloading and sharing music (Coats, 2010). Teenagers, campuses, and even celebrities have been sued. In 2007 alone, the RIAA sued 35,000 individuals for illegally downloading copyrighted music. Moreover, the fines, even when lower than suggested, can be substantial — up to $150,000 per violation. As one example, in 2009 a court issued a judgment of $22,500 against a doctoral student for each of 30 downloads, for a total of $675,000. A district court subsequently reduced that amount by 90%, but the First Circuit Court of Appeals reinstated the original judgment in September 2011 (Smith, 2011).

2. **Content owners employ digital rights management (DRM) technologies to prevent illegal copying.** DRM technologies ‘include and at times combine encryption, digital signatures, “fingerprinting,” serial copy management, watermarking (also known as steganography),’ and other mechanisms (Fayenson, 1999, p. 9). Such restrictions enabled Apple to establish a virtual monopoly in the field of downloaded music by prohibiting users from playing their acquired music on any media other than an iPod. These restrictions also collided head-on with the widely held consumer belief that ‘copyright laws should not infringe on an individual’s access to music that they have legally purchased’ (cited in Fischer, 2004, p. 101). Technologically savvy users retaliated by removing the DRM mechanisms via reverse engineering in an ongoing game of ‘cat and mouse.’ ‘As quickly as a producer figures out a way to encrypt a DVD or software program to prevent duplication,’ observed one critic, ‘some hacker in Seattle, Reykjavik, or Manila figures a way around it’ (Clement, 2003, p. 3).

Regardless of who wins which copyright battles, these ongoing conflicts generated irresistible pressures to formulate new public policies to clarify copyright and fair use for the digital age. The key legislative response was the Copyright Act of 1976, the first major modification of U.S. copyright law since 1909. Section 102 (b) explicitly reaffirmed that copyright protects the expression of ideas, but not the ideas themselves. Section 107 identified the four criteria for fair use (see above). It also extended the duration of copyright protection—for works created on or after January 1, 1978—to the author’s lifetime plus 50 years. Given the increasing lifespan in the developed world, a copyright awarded to a young author or composer could easily extend beyond a century—while benefiting economic enterprises that acquire large bundles of copyrights in ways not previously imagined. As a result, this section of the law was perceived as a victory for copyright owners.

**Global Piracy**

As the United States adopted strategies to address domestic piracy, global piracy mushroomed into a major international issue with both economic and political attributes. According to the Organization for Economic Cooperation and Development (OECD), in 2009 commerce in pirated and counterfeit products totaled roughly $250 billion, or nearly 2% of annualized global trade for that year (cited in WIPO, ‘Counterfeiting’). Efforts to deal with international piracy have had significant implications for current U.S. copyright policies.

In 1886, the Berne Convention for the Protection of Literary and Artistic Works established basic rules and definitions concerning authors’ rights. As with U.S. copyright laws, the Berne Convention was periodically revisited and updated. The most significant revision was the Paris Act of 1971, which specified six ‘exclusive rights’ of authors (cited in UKCS Fact Sheet P-08):
HOW COPYRIGHT THEORY AFFECTS PRACTICES

1. ‘The right to authorise translations of the work.

2. The exclusive right to reproduce the work, though some provisions are made under national laws which typically allow limited private and educational use without infringement.

3. The right to authorise public performance or broadcast, and the communication of broadcasts and public performances.

4. The right to authorize arrangements or other types of adaptation to the work.

5. Recitation of the work, (or of a translation of the work).

6. The exclusive right to adapt or alter the work.’

Responsibility for implementing the standards imposed by the Berne Convention lies with the World Intellectual Property Organization (WIPO), which became a specialized agency within the United Nations in 1974 (WIPO Treaties).

Significantly, the United States chose not to sign the Berne Convention. U.S. refusal to participate in global copyright efforts reflected a basic ideological difference between the United States and Europe. For several centuries, the American Colonies (and later, the U.S.) emphasized users’ rights. In fact, the Republic frequently ‘borrowed’ literature, technologies, and similar intellectual products from Britain and other European nations (e.g., colonial publishers rarely paid European authors or their publishing houses). In contrast, wealthy European nations prioritized authors’ rights. This dynamic is often reprised on the world’s stage—i.e., less-wealthy nations frequently resort to global piracy to develop their economies, and individuals view illegal file sharing as a form of ‘economic fairness’ (Mitchell, 1994; Shelfer, 2002) or political activism (e.g., WikiLeaks). Over time, however, the United States transitioned from a copyright importer to a major copyright exporter. Consequently, the official U.S. position became closer to the European model—i.e., a victim rather than a victimizer (Vaidhyanathan, 2001, pp. 160–61). Reflecting this transition, the U.S. officially joined the Berne Convention in 1989 (U.S. Court of Appeals, 2010).

Uruguay Round

The 1990s witnessed two major global initiatives to protect intellectual property that ultimately reshaped U.S. copyright policies. The first occurred in 1994 when the Uruguay Round transformed the General Agreement on Tariffs and Trade (GATT) into the World Trade Organization (WTO). All WTO members—including the United States—were required to become signatories to the Agreement on Trade-Related Aspects of Intellectual Property Rights, or TRIPS, which essentially incorporated the standards of the Berne Convention. Two years later, WIPO promulgated the WIPO Copyright Treaty (WCT), which required all contracting parties to adhere to standards of the 1971 Paris Act (WIPO, 1996). One highly controversial clause required all members to ‘provide legal remedies against the circumvention of technological measures (e.g., encryption)’ designed to protect copyright. This language appeared to prohibit all forms of reverse engineering. Such a reading would place the more liberal provisions of the 1992 Sega ruling in violation of international law (Mitchell, 1994; WTO, 2010).

The WTC and related international actions created pressures for the United States to conform its practices to the generally more-restrictive European standards. The first significant response was the passage of the Uruguay Round Agreements Act (URAA) in 1994. The URAA essentially restored copyright protection to foreign works that formerly were in the public domain. Included here were Picasso’s Guernica; films by Fellini, Godard, and Hitchcock; writings by H.G. Wells; and musical compositions by Shostakovich, Prokofiev, and Stravinsky. Fair use advocates, including the Electronic Frontier
Foundation (discussed below) and the CIS (see above), have challenged the constitutionality of the URAA. They formalized their opposition into a case — *Golan v. Holder* — for which the Supreme Court began to hear arguments on October 5, 2011 (Decherney, 2011; Falzone, 2011; Liptak, 2011; Supreme Court of the United States, 2011).

**Digital Millennium Copyright Act**

The most significant move to align U.S. policy with international standards was the passage of the Digital Millennium Copyright Act (DMCA), signed by President Bill Clinton on October 28, 1998. (To dispel any doubts as to its objectives, this law was also titled the ‘WIPO Copyright and Performances and Phonograms Treaties Implementation Act of 1998.’) The DMCA revised — and essentially doubled the size of — the existing copyright law (Fisher, 93).

The most consequential and also the most controversial provisions of the DMCA were the anti-circumvention regulations contained in Section 12. Section 1201 made it illegal to ‘manufacture, import, offer to the public, provide, or otherwise traffic in any technology, product, service, device, component, or part thereof’ designed to circumvent any security measure ‘that effectively protects a right of a copyright owner.’ Section 1202 outlawed any activities designed to intentionally remove or alter any copyright management information. Significantly, this prohibition applied to all such efforts, regardless of whether they technically constitute copyright infringement.

In essence, the DMCA rendered the reasons for reverse engineering largely irrelevant. Manufacturing or purchasing anti-encryption software became illegal, regardless of how the technology was (or wasn’t) used. These provisions represented a clear victory for content providers. They also appeared to nullify the key elements of both the *Sega* and *Sony* decisions. To cite one critic: ‘The DMCA makes the kind of reverse engineering commonplace in most industries illegal in copyright works’ (Doctorow, p. 72).

The law did allow certain exceptions based on fair use. However, these exceptions are ‘fairly cumbersome and, at times, difficult to wade through’ (Fayenson, 1999, p. 10). As a concession to fair use advocates, the law also authorized the Librarian of Congress to conduct periodic reviews to determine whether certain categories of copyrighted materials should be exempt from these restrictions. Overall, however, the DMCA had reformulated copyright protection in a manner that dramatically shifted the balance of power in the direction of content owners.

The DMCA also redefined the enforcement process. Prior to the passage of the act, a copyright owner who claimed infringement was required to submit a complaint to a court, which would then determine whether the accused party had actually violated the owner’s rights. This approach extended significant protection to the content providers (e.g., ISPs). The DMCA modified the enforcement process as follows:

- Under the DMCA, when a content owner such as the RIAA suspects that its information is being used inappropriately, it issues a takedown notice to the provider. The provider must remove the offending material and give the offending party an opportunity to submit a ‘counter-notice’ disputing the charge (Coats, Lerner, and Krause, 2010; Fayenson, 1999).

- Thus, while awarding ‘subpoena-like powers’ to copyright owners, the DMCA effectively placed the responsibility for combating copyright infringement on the shoulders of Internet service providers.

**Example 6**. The DCMA prohibited consumers who purchased music from iTunes from overriding encryptions or other DRM technologies that prevented them from copying the music to another format, e.g., CD or MP3. Violators were subject to significant civil and, in some circumstances, criminal penalties (Electronic Frontier Foundation, 2010).
(ISPs) and online service providers (OSPs) (Van Horn, 2002). Significantly, the latter group includes libraries and universities. In one instance, a copyright holder (RIAA) actually sued an ISP (Verizon) to obtain the identity of a customer it accused of illegally sharing music online (Sisario, 2011).

- The act provides a ‘safe harbor’ for providers—i.e., they cannot be held liable by owners for infringement as long as they are not consciously aware that people are using their services illegally (Coats, Lerner, and Krause, 2010). To qualify for safe harbor status, however, the ISP must have implemented a reasonable termination policy.

- For several years there was no uniform strategy to enforce the DMCA regulations. Finally, in July 2011, major ISPs — including Verizon, AT&T, Comcast, Cablevision, and Time Warner Cable — reached an agreement with media companies such as the MPAA and the RIAA on a standard procedure. Essentially, the ISP will send a series of up to six online notifications to users accused of illegally downloading copyrighted materials. Users who fail to respond will be subject to increasingly severe penalties culminating in dramatic reductions in their Internet connection speeds and blockage of their Web-browsing capabilities. In addition, ISPs retain the right to terminate service to users who consistently violate their contract terms (Sisario, 2011).

Copyright Term Extension Act

In the same year that Congress enacted the DMCA, it also passed the Copyright Term Extension Act, also referred to as the Sonny Bono CTEA, because Bono, a performer–turned-congressman, was pushing for its passage at the time of his death. The CTEA increased the copyright protection period to the author’s lifetime plus 70 years. Advocates for public access challenged this law in the courts, arguing that empowering the Congress to continually extend the duration of copyright protections violates the constitutional stipulation that copyrights be awarded only for a limited amount of time. In 2003, however, the Supreme Court ruled in *Eldred v. Ashcroft* that Congress possesses the constitutional right to extend copyrights (*Eldred v. Ashcroft*, 2003; Lessig, 2004b).

Beyond the DMCA

Fair use proponents quickly launched an assault on the restrictive aspects of the DMCA. The ALA website, for example, issued the following message: ‘The doctrine of “fair use” has never more been threatened than it is now’ (American Library Association, DMCA). Law professor Pamela Samuelson, while acknowledging that ‘the Internet presents new challenges to copyright,’ nevertheless asserted that ‘giving private companies such strong control over the flow of information in society is not the right answer’ (1998, p. C03). Lawrence Lessig of the CIS expressed concern that the DMCA would severely restrict the democratic potential of the digital age by awarding ‘the monopolists of culture’ — exemplified by the RIAA — excessive control over copyrighted material (2004a, p. 184). The result, he contended, will be a ‘permission culture’ as opposed to a ‘free culture.’ Finally, writing in *Phi Beta Kappan*, education professor Royal Van Horn (2002) characterized the DMCA and the CTEA, along with the Patriot Act and other laws limiting Internet access, as components of a broader movement to limit Americans’ constitutional rights, which intensified subsequent to 9/11.

The movement to expand copyright protection catalyzed the formation of activist groups that support the unimpeded flow of information. Among the most prominent are the Electronic Frontier Foundation (EFF), co-founded by John Perry Barlow of the Grateful Dead, and the Digital Future Coalition (DFC). Not surprisingly, these and other groups have criticized and challenged the anti-circumvention provisions of the DMCA. For example, a March 2010 EFF report asserted: ‘The anti-circumvention provisions have been used to stifle a wide array of legitimate activities, rather than to stop copyright infringement.’ The report identified specific threats posed by the act’s stringent regulations, including:
1. Jeopardizing free speech
2. Impeding scientific research
3. Severely restricting fair use rights
4. Stifling competition, thereby curtailing innovation (EFF, 2010).

Significantly, since 1998, some of the more liberal interpretations of fair use have been restored. In 2008, for example, a U.S. district court ruled in favor of Stephanie Lenz, declaring ‘the unnecessary removal of non-infringing material causes significant injury to the public’ (*Lenz v. Universal Music Corp.*). Specifically, the court ruled that content owners must consider fair use when they issue takedown notices. Significantly, this stipulation was not included in the DMCA. Then, in July 2010, Librarian of Congress James H. Billington ruled that the reverse engineering of smart phones to download additional applications — a process known as ‘jailbreaking’ — does not constitute a copyright infringement under the DMCA (U.S. Copyright Office, 2010). This ruling came in response to a complaint submitted by the EFF. (Recall that the DMCA empowered the Librarian of Congress to rule on exceptions to DRM protections.)

The *Lenz* and LOC decisions demonstrated that the fair use movement has survived — and perhaps even been invigorated by — the passage of the DMCA. Grassroots discontent with expanding copyright protection, coupled with the negative publicity generated by corporate efforts to limit users’ rights and prosecute violators, seems to have penetrated the upper echelons of the information industries (see Example 7).

**Example 7a.** In 2009, Apple announced it was removing DRM restrictions from its digital music because they discourage sales (AppleInsider Staff, 2007; Cohen, 2009).

**Example 7b.** In 2010, the RIAA announced it would replace its litigation-based strategy for combating copyright infringement with a more cooperative approach (Coats, Lerner, and Krause, 2010).

Today, the copyright debate is unfolding in a bitter partisan political environment. Underlying many of the specific arguments is the ongoing question of whether copyrights and patents essentially encourage or discourage creativity and innovation. One of the leading proponents of maintaining copyright protection is Richard Epstein, a law professor at the University of Chicago who is associated with the Hoover Institution. Epstein is a leading figure in the property rights movement (PRM), a generally conservative/libertarian movement that opposes what they perceive as a governmental assault on traditional property rights. This philosophy fits comfortably within the larger framework of antigovernment sentiments that are prominent in contemporary U.S. politics. Epstein argues that copyright protection should cover the full range of inventions and writing. Similarly, patents should apply ‘to “anything under the sun” that results from human labor and imagination’ (2010, p. 484). Although Epstein acknowledges that intellectual property differs from physical property in many ways, he maintains that owners of both forms of property are entitled to some form of protection. He further posits that the state possesses the authority to create a process to protect creators’ rights. Epstein contends that ‘offering limited protection to creators through patents and copyrights will introduce innovation that will redound to the benefit of the outsiders in the long run’ (p. 482).

In contrast, critics of the DMCA reject the argument that intellectual property can and should be subject to the same regulations as physical property. For example, Peter Menell of the Berkeley Center for Law & Technology asserts: ‘Property is not a monolithic concept, and its treatment varies significantly across classes of resources’ (2007, p. 721). Thus, the rules protecting homes and automobiles, for example, cannot be applied to cyberspace. ‘Intellectual property,’ he contends, ‘has never fit the real property mold particularly well’ (p. 753). Menell concludes that a strictly limited degree of copyright protection is appropriate. However, any system designed to protect intellectual property must be dynamic so it can adjust to the continual change.
that defines the digital era. Because traditional physical property rights as defined by the PRM are static, he concludes, they cannot serve as the foundation for such a system.

Despite the obvious divergence in the views of the property rights movement and the digital rights movement, both schools of thought accept protecting intellectual property as a basis for copyright. According to Vaidhyanathan, ‘The republican roots of copyright are almost forgotten in public discussions of copyright and virtually absent from the concerns of policy makers’ (2001, p. 37). Based on our research, we believe that technology has triggered new forms of ‘populism’ associated with socially acceptable infringements of copyrights. As a result, the late twentieth and early twenty-first centuries have witnessed numerous instances of ‘republican’ protests against current copyright laws and substantive adjustments of copyright owners’ enforcement practices. The pendulum may be swinging back toward advocating some notion of information as a ‘public good,’ or at least reconsidering the ‘republican’ theories that initially underlay copyright laws and practices.

CONCLUSIONS AND RECOMMENDATIONS

Copyright is a vitally important issue for librarians and other professionals who work with information. As technology continues to evolve, disputes between stakeholders will intensify. We believe that various elements of bundled copy privileges are likely to reflect (and require) more than one set of rules and regulations, partly because the private use of modern technologies has transformed society’s basic understanding of both property and ownership and is now transforming society’s expectations as well.

Librarians and other information professionals are not simply caught between their need to preserve the economic sustainability of content providers and their mission to ‘inform’ the communities they serve. They are also agents of social change. For this reason, we recommend the following professional management practices:

1. **Librarians are cognizant of and adhere to the current regulations concerning both copying and reverse engineering.**

   One of the ALA’s fundamental principles, as expressed in its Code of Ethics, asserts: ‘We respect intellectual property rights and advocate balance between the interests of information users and rights holders’ (ALA website). This principle commits librarians to (1) accepting the validity of intellectual property rights and (2) acknowledging the library’s (and their own) responsibilities to abide by copyright laws.

2. **Libraries monitor the changing expectations of creators, owners, and users of copyrighted information** to assure that both copyrights and copy privileges are protected by library staff.

   As Example 3 (above) indicates, the definition of the *legal* acquisition and use of copyrighted information is problematic. Ubiquitous global noncompliance is forcing societies and governments to rebalance numerous competing interests. Reflecting this reality, the ALA currently devotes a segment of its website exclusively to the DMCA. Because libraries qualify as OSPs, they (1) may become the target of takedown notices by content owners as stipulated by the DMCA and (2) can be compelled to turn over the names of offenders to the copyright owners. Librarians therefore must devise strategies that protect patron privacy without aiding and abetting lawbreakers.

3. **Patron privacy is protected—when it should be.**

   As OSPs, libraries are positioned to benefit from the safe harbor provisions of the DMCA. Recall, however, that the DMCA protects OSPs only if they facilitate (aid and abet) copyright infringement
without being aware that they are doing so. Neither theft nor sabotage is an acceptable professional practice. Moreover, safe harbor protection does not apply when libraries post the copy-protected materials themselves.

- Librarians do not break the law. They avoid posting materials that infringe on copyrights; they remain alert for indications that they have posted such materials (Stanco, 2000); and they respond expeditiously if/when they are informed by copyright owners of any infringements. They review guidelines and updates provided by information associations, attend copyright briefings at conferences, participate in IP workshops, and make certain that support staff are briefed and provided with appropriate in-service training. They also track library resource-sharing activities such as interlibrary loan (ILL) and content placed on reserve, both inside and outside e-reserves systems (e.g., using paper and pencil, office productivity software such as Excel™ spreadsheets, and expert systems such as ILLIAD and ARES). In addition, they pay the appropriate fees.

- Librarians do not aid and abet lawbreakers. They post fair use policy statements at selected service points and on library websites. They also post sets of frequently asked questions (FAQs; e.g., a quick Google™ search yields content uploaded by the University of Oregon Libraries), fair use checklists (e.g., Columbia University Libraries Copyright Advisory Office), and other educational material (e.g., the Stanford Universities Libraries fair use charts and tools).

Librarians also reach out to inform the communities they serve, posting copyright and fair use information in online social networks and learning communities. For example, the North Georgia College and State University Library Technology Center posted a YouTube™ video on April 11, 2011, which can be accessed at http://www.youtube.com/watch?v=bc__AiZf4NI. Librarians may also publish op-ed articles in newsletters and newspapers and deliver high-quality, in-service training, including workshops for teachers and members of civic and professional associations, to ensure that copyrights and fair use privileges are both protected.

Establishing and maintaining a balance between conflicting objectives represents one of the most compelling challenges faced by twenty-first century librarians and other information professionals. Whether the democratic nature of cyberspace will reverse recent trends concerning intellectual property is a vital question for the digital age. Librarians are entrusted to appropriately manage the authorized use of copyrighted information. It follows, then, that libraries should inform other stakeholders of their rights and responsibilities. In addition, our professional associations should continue to play an active role in assuring that copyrights evolve both to meet changing opportunities and to mitigate the emerging threats of the digital age.
BIBLIOGRAPHY

American Geophysical Union v. Texaco Inc., 60 F.3d 913 (2nd Cir. 1994).


Coats, W.S., Lerner, L.L., & Krause, E. (2010, January). Preventing illegal sharing of music online: The DMCA, litigation, and a new, graduated approach. Journal of Internet Law 13(7), 3–7. Available at http://web.ebscohost.com/ehost/results?vid=3&hid=13&sid=86351d8c-4823-45c8-970c-2913f7c63631%40sessionmgr113&bquery=%28dmca%29&bdata=JmRiPWY1aCZkYj1ic2gmZGI9bmZoJmRiPWVyaWMmZGI9bW1oaCZkYj1mdW5rJmRiPWh4aCZkYj1id2gmZGI9dWxoaJmRiPW1laCZkYj1sZmgmZGl9c2NoJmRiPXRxmaCZkYj04Z2gmZGI9XYBoJmRiPW1laCZkYj1raGgmZGI9aGNoJmRiPWx4aCZkYj10dGgmY2xpMD1TTyZjbHYwPWpvdXJuYWwrb2YraW50ZXJuZXQrbGF3JmNsaTE9RFQxJmNsdjE9MiAxMDAxLTlwMTAwMiZ0eXBiptAmc2l0ZTlaG9zdC1saXZl.


U.S Copyright Office. (2010, July 28). Rulemaking on exemptions from prohibition on circumvention of technological measures to control access to copyrighted works. Available at http://www.copyright.gov/1201/.


ABOUT THE AUTHORS

Robert J. Weiss obtained his doctorate in history from New York University and has just received his MLS from St. John’s University in New York. Dr. Weiss is the author of “We Want Jobs”: The History of Affirmative Action (Garland Publishing [Taylor & Francis], 1997) as well as several brief monographs. He also served as associate editor of The Papers of William Livingston, published by Rutgers University Press. In addition, Dr. Weiss is an experienced college textbook editor, having worked for many of the nation’s most prominent publishing houses across a broad spectrum of disciplines. He has taught on both the high school and college levels, and, as an undergraduate, he supervised a tutoring program for Upward Bound high school students.

Katherine M. Shelfer is an Associate Professor at St. John's University. She received all three of her degrees from The Florida State University, and she has pursued advanced graduate work in the humanities with a focus on medieval and Renaissance commerce. In addition, Dr. Shelfer has held certification/licensure in teaching, securities, all forms of insurance, and real estate. She also has human resources experience in two Fortune 100 companies, in addition to extensive library-related experience. Her publications cover the selection, evaluation, management, and use of electronic and print resources for business and competitive intelligence, as well as the integration of knowledge management and competitive intelligence. She was voted a Fellow of the Society of Competitive Intelligence Professionals and is currently serving as Editor, IAFIE News.
RFID Materials Circulation and Handling: A Model for Improving Customer Service

By

Christine McDonald

Abstract: In the mid-2000s, at ALA and PLA conferences, all indications were that the Elvis impersonator at Checkpoint™ was finally showing a product that had most of the bugs worked out, was more reasonable in cost with the promise and trend of individual RFID tag costs coming down each year. With other RFID vendors improving products and with many more companies offering competitive ideas and pricing in the RFID arena, my library, Crandall Public Library in Glens Falls, New York where I serve as director, started the formal process of exploring RFID in 2004. I had watched the progress of RFID for more than 15 years, convinced that the technology would eventually improve and transform how we do business in the circulation departments of public libraries. From 1999 to 2005, I also served on the IFLA [International Federation of Library Associations] Services to MultiCultural Populations Committee. At IFLA conferences and on visits to foreign libraries, the use of materials handling systems was in use in many European libraries and exceeded what is being done in the U.S. in some cases. One sorting approach uses a robotic system to take materials from sorting bins and place them in number order on carts!

Introduction: Why do It?

It costs a Library considerable staff time, board time, and money to implement RFID and these are large hurdles that must be navigated. Thus, there are three major opportunities to implement RFID: During a new building process, during a major renovation or as part of a remodeling project. At Crandall Public Library, we started working on a major renovation project in 1996 when we did a space study with Nolan Lushington.

Christine McDonald is Director of the Crandall Library and can be reached at: mcdonald@crandalllibrary.org
After an unsuccessful bond vote in 2000, we had a successful bond vote for $12.875 million in 2005. And, we raised over $6 million to fund an $18.8 million project of new construction and a complete remodel of an historic 1930s Charles A. Platt building. During the planning stages, my board president, a retired GE engineer and I read the current literature on RFID and I continued to see demos at PLA and ALA. For many reasons aside from financial, I was convinced RFID self check and a materials handling system for returns was the best solution for the long term interests of the library and the community. What impressed me and my board president as we reviewed RFID literature were the number of municipalities that were advising or insisting that their public libraries consider RFID as a means to control staffing costs and to increase efficiency of operation. RFID promises to get materials checked out fast with minimal staff intervention.

When the bids were opened for our project in April 2007, and they were about $1 million under estimates, I turned to my board president and said, “Now we can get RFID.” He agreed it was a priority. We also knew that funding from the NYS Public Library Construction Grants would be available to assist us in implementing RFID. We did apply for and receive a grant from this program for a portion of the project in 2008-09 as did the Saratoga Springs Public Library in 2011 when they added RFID self check, but not a sorting system. The NYS Public Library Construction Grants are a viable way for libraries to consider RFID and sorting systems. The grants cover 50% of allowable costs, leaving libraries the task of raising the other 50%.

However, as a result of adding RFID after we opened the bids meant that the architect had to redesign the customer service area to accommodate self check units, DVD unlocking stations, payment machines and a sorter room for the conveyor belt sorter and bins. In essence, we were doing a renovation of a newly designed building. We were fortunate that our architect, Ann Beha Architects from Boston, MA were amenable to making this eleventh hour change. Not only did they redesign the area, but the architect we were working with, Steve Gerrard, came with us on a field trip to a Massachusetts library in his neighborhood, the Norfolk Public Library to view an RFID system in operation.

How we chose the RFID vendor would be an article in itself; however, the process was comprehensive, time consuming, and time limited since we had to meet deadlines from the architect and construction manager; however, the RFID system Crandall Public Library selected was a proposal from Polaris/ITG after the field was narrowed down from eight to three vendor proposals.

**RFID became a necessity, not a luxury.**

Knowing we would increase circulation by 40% or more with a newly renovated and expanded facility, which translated into adding five new staff members, was our main reason we felt that adding RFID was cost justifiable. When the Saratoga Springs Public Library in 1995 and Clifton Park-Half Moon Public Library in 2005 opened their...
doors after building completely new buildings, their circulation rose over 40%. I called many directors of other libraries of similar size like Louise Schaper of Fayetteville, Arkansas and heard the same story. I also phoned directors and staff of libraries that opened with new RFID systems in place to find out their experiences. The advice was to prepare to staff the circulation department for a 40-50% increase. When Clifton Park opened, they had lines out the door from the circulation desk. At Crandall, we analyzed the data and determined we needed at least five new full-time staff members to handle circulation after we opened the new building plus the processing of materials with an increased materials budget. Our circulation staff is assigned to both circulation duties and processing of materials. We knew we couldn’t get a budget passed if we proposed adding five new staff members. As it was, our budget for 2009 was passed with a 19% increase to handle the increase in operating costs with a newly renovated and expanded building. Five new circulation staff added to that budget would have brought it to the 28-30% level. We knew that was impossible to achieve such an increase in the economic climate of 2008.

If a library chooses to include a materials sorting system to check in materials as we do at Crandall Public Library, this is adds the benefits of eliminating staff intervention at the point of return. The value of the sorter can’t be minimized. It eliminates huge amounts of work that technology can do better than staff and it saves staff time which can be better spent assisting customers. Rather than check in materials manually, re-sensitize for security and hand sort by type of material and location, the sorter does all that work instantly as the item moves down the sorter conveyer belt. The larger the number of bins selected, the more flexibility the library has for getting materials back on the shelves fast. At Crandall, we have a seven-sort bin system by type of material and floor. Library Pages take the materials from the bins to carts and wheel them out to be re-shelved. We have almost eliminated “holding” shelves using this system. Also eliminated were repetitive motion injuries with RFID sorting and check out. Prior to using RFID, many circulation staff members wore braces for their hands because of using a light wand for hours to check in and check out materials. One person had carpal tunnel surgery.

We opted to place a credit card payment device on each self check unit to assist customers in paying fines. And we purchased coin/bill payment machines for two self check areas, the main Customer Service Area and the library’s park entrance. Later we added another coin/bill payment machine in the Children’s Department. While the machines don’t eliminate all money transactions using staff, these transactions dropped significantly from pre-RFID days. This helps accomplish one of our main goals with RFID of using technology to free up staff time for more meaningful interactions with customers.

RFID Security challenged our concerns for fast customer service vs. theft with AV items

Security with RFID is both simpler and more complicated than traditional security systems. We had been using 3M™ security strips which had to be inserted into each type of library material in addition to bar coding each
item. As part of the RFID selection process, we evaluated RFID security systems. Books are easy to secure both with RFID and traditional security products. But, with an RFID system, the same RFID device in each book serves both as the security strip and the means to check out the item. RFID security tags are much easier to insert than the strips we were using. But, we found to our dismay that security for DVDs, one of our collections that, for its size, circulates more intensively than any other material type wasn’t simple. Many RFID vendors sell round “donut” strips specially made for DVDs and CDs and require a booster on the security system to read this special tag. They are pricey too. Equally costly are DVD locking cases. Our choice between locking cases and “donut” security strips resulted in our decision to go with unlocking cases. The success rate using the “donut” strips wasn’t high enough for a collection that gets maximum use for a relatively small collection size. We needed to maintain the integrity of the collection by not losing huge numbers of DVDs to theft. To help minimize the cost of the locking cases, we noticed that many libraries were doing away with the cases and selling them online. We were able to purchase hundreds of these cases for a fraction of their cost. And the Friends of Crandall Public Library also purchased some of the cases for the library. We purchased five unlocking stations, locating two back to back opposite the self checks in the customer service area. Unlocking stations are located wherever self check units are located.

Expectation vs. Reality: How the Circulation Department Transformed Into the Customer Service Department

During the process of selection of our RFID system, I spoke with staff from many libraries using RFID. Success rates of RFID use by customers at various libraries we polled ranged from less than 20% to 90%. We decided that 80% use for RFID self check would be our preliminary goal. There’s a learning curve in using the system and we anticipated that customers and staff would need time to make the transition too. RFID changed our way of doing business, gave us new insights into how to deliver better customer service and challenged our thinking about the role of staff. We started questioning our processes. Before we opened our doors to the renovated and expanded facility, we had renamed the Circulation area to the Customer Service Area since so many customer functions were handled by the staff in this area.

After six months, we found that only 65% of check outs were with the RFID self check units. I knew something was wrong and started to investigate. What I learned through meeting with customer service staff was that while they found we couldn’t handle the volume of traffic without RFID check out and check-in, staff was concerned about giving good customer service. In the past that meant checking out materials as quickly as possible and interacting with customers. Since many customers were learning a new system, staff felt it was faster to check out those customers who were unsure of how to use RFID thereby fulfilling their role in providing good customer service.
Some staff felt they weren’t serving customers if they weren’t checking their materials out for them. I started having weekly meetings with the customer service staff to hear all their concerns. When you implement RFID, it’s a sea change for everyone. Staff didn’t yet see that teaching customers how to use RFID self check was interacting with customers at a higher level because they were teaching a skill. The teaching role is one which library staff has been doing throughout our history. During the discussions with staff, we acknowledged that the process of finishing the building, passing a 19% increase in the budget and opening the building on the target date meant that the practical issues that come with change weren’t addressed as comprehensively as we had hoped. But we were only into the first six months of RFID operation and we were all learning from the new system and from each other. We explained to staff that checking out materials with a light wand was not as complex as teaching a customer how to check out materials, unlock DVD cases and pay fines automatically using the RFID self check stations which accept coins, credit cards and bills up to ten dollars. As staff was encouraged to understand their role as teachers with customers, the RFID use increased to 75% in less than two months. Our staff gave us meaningful feedback that resulted in all of us understanding the new service roles brought about by the change to RFID circulation. Currently RFID self check use 81%. Our goal is to get to 90% by 2013.

In addition to the teaching role with RFID check out, our Reference staff mentored the Customer Service Staff in learning the functions of the OPACs that customers would benefit from using. Customer Service Staff now routinely explain to customers how to renew books online, how to place reserves, how to find materials, how to see their fines and other OPAC functions.

Results

Looking at 2009 statistics, the first full year of operation with RFID, circulation increased 47% which was beyond what we expected. Staff agreed that without RFID self check and sorting, keeping up with such an increase would have been impossible without adding staff. Lines would have been a reality. Instead, RFID self check allowed us to keep up with the demand for circulation services.

When we first opened for business in December 2008, we had six self check units. Within a year, we added an additional unit in the main Customer Service Area because we were getting lines during peak check out times. The third unit in this area eliminated lines to this day. We also added a coin/bill payment machine in the Children’s Department. To keep transactions requiring staff intervention to a minimum, we analyzed all transactions coming across the main Customer Service desk and discovered that a fair amount of activity was for payment of children’s fines. Customers were annoyed when they had to go down one level to pay fines. It was something we overlooked in the planning but it was quickly resolved.

When we first opened our doors with our fully automated self check system, there was resistance by customers and staff. This is normal when any huge change is initiated. Customers were unsure if they could figure out how the system worked. It didn’t take long for children to take to the new check out and check in system. Children liked how much fun it is to check out materials and many taught their parents how to use the system. Many adults found out quickly that the check out process goes fast, literally within seconds, without any lines or waiting for someone to help. They found that checking out a book or DVD was easier than checking out a grapefruit or tomato at the supermarket self check-out. The architects designed the sorter area with a 90 degree glass wall allowing customers to watch as materials move down the conveyer belt and deposited in the correct sorting bin. Children of all ages love to watch the sorter in action. When materials go through the sorter, the item is immediately deleted from a customer’s record. There’s no waiting for a staff person to check in the materials. For a customer who is returning DVDs (which have a title limit), once the DVD goes through the sorter, the customer is immediately
able to take out more DVDs. This is a real plus in providing good customer service. It’s fast, efficient and makes customers happy. When we did our Snapshot NY 2010 survey, we only got one comment about self check, “It’s easy to use. I like self check out.” Because customers find self check fast and simple, it has become part of the way Crandall Public Library normally does business. As technology continues to offer ways to offer better customer service, all of us as library staff are better able to provide more complex services with our current decreasing resources. I can’t imagine working in a public library today without RFID technology. It frees up staff to do important customer service work and empowers staff to learn new skills.
DIGITIZING EVERYTHING?

Part II: PILOTING METADATA CREATION

By Kimmy Szeto

Abstract: The Queens Borough Public Library launched the Digital Assets Management System in 2009 with the ultimate goal of digitizing all the holdings in the library's Archives. A pilot project was initiated in the course of designing the metadata creation policy and cataloging workflow. This paper discusses key policy and design elements such as imaging requirements, legacy data migration, metadata schemes, data formats, file naming, and controlled vocabulary, and presents sample data processing scripts, VRA Core 4.0 metadata records, and transformations to HTML and KML documents.

Introduction

The Queens Borough Public Library (QBPL) is nearing the public unveiling of a web site for its digitized archival collections dedicated to the history of Long Island. The driving force behind this effort is the Digital Assets Management System (DAMS), which was initiated by its current Project Manager, and was formally launched in 2009 after 3 years of planning (Szeto, 2011). I joined the digital program as its metadata librarian at that time, and participated in developing the program’s metadata policy, workflow design, and implementation. In this role, I frequently drew ideas from current literature on digital imaging implementation, and at the same time invented solutions for site-specific problems. The detailed account of the metadata creation process for the digitization program’s pilot project in this paper provides a practical perspective of our course of action.

Literature on Implementing Digitization Programs


Kimmy Szeto is Assistant Librarian at the Maritime College, State University of New York. He heads cataloging and metadata creation at the Stephen B. Luce Library and Archives, and manages the Federal Deposit Library Program at the library. His email is kszeto@sunymaritime.edu.
Novara (2010) discusses the digital imaging and archival practice at the University of Maryland, with emphasis on adapting workflow to the changing demands researchers make when it comes to digital technology. Schmidt, et al. (2011) detail the survey, evaluation, and planning carried out by the digital curation team at Michigan State University. Their findings address how one institution makes decisions on metadata schemas, controlled vocabularies, and digital storage solutions in response to various local practices. Fox (2008) focuses on descriptive metadata for digital images and carefully examines the VRA Core schema and its role in capturing metadata for cultural heritage materials. Colati and Colati (2011a, 2011b) impart many words of wisdom as they chronicle the effort of a fictitious academic librarian who accidentally and reluctantly found himself tasked with establishing a digital contents management program.

Deepening the level of detail that such sources have begun to provide, this paper focuses on solving metadata problems during the first implementation of a digitization program. Presenting this experience will help other libraries overcome some of their implementation hurdles.

**Project Background**

**The Digital Assets Management System and the Archives**

In 2006, the Queens Borough Public Library established its Digital Initiative to digitize the contents of its archives. After the initial feasibility study and organizational structuring, the Digital Assets Management System was set up as a subdivision of the Archives. By 2009, it had established its mission and collection development policy, had been staffed with 3 FTEs, and had had hardware and software procurement funding approved. The scope of this program was ambitious: it would digitize all the contents of the library’s Archives and the library’s institutional records, as well as collect all born digital contents emanating from the library. This presented a curation predicament similar to the one found in Schmidt, et al. (2011)—each material format in the Archives had its own set of metadata, and each individual agency produced its own electronic documents.

The holdings of the Archives consist of roughly 3,600 monographs, 2,500 cubic feet of manuscripts (in about 100 collections), 4,500 maps and broadsides, 105,000 photographs and postcards (in about 50 collections), 425 feet of vertical files (about 2,500 files), and 9,000 reels of microfilm (about 200 titles) (QBPL, 2011). All the monographs, vertical files, and titles on microfilm have MARC records and are searchable on the library’s OPAC. In 2009, the cataloging department had just begun to catalog all the maps and broadsides systematically, also in MARC. Manuscripts had all been accessioned by librarians in the Archives, who had also created finding aids for each collection. These finding aids were in the process of being encoded in EAD at a steady pace by volunteers. Earlier in the mid-1990s, most photographs had been described at the item level and digitized at low resolution. Ever since then, these images have been available for searching and viewing at a computer terminal using the ApplicationXtender database software by EMC.

With the procurement process ongoing, it would take another 6 to 9 months before the new imaging system materialized; at the same time, the software company contracted for the digital contents management software was experiencing production delays. In the meantime, the digital imaging program, firmly established and fully staffed, found itself without the equipment to do the work. However, there was one area to explore. The Archives Division owned a high-resolution flat-bed scanner, and offered a fee-based scanning service, where library patrons could request high-resolution scans of images discovered with the aging ApplicationXtender.

We saw a convergence of factors: a high-quality scanner, an outdated database, and a need to improve access. It made sense to digitize photographs systematically using the existing flat-bed scanner, and delve into metadata creation—developing policies, creating workflow, and migrating legacy data—through a pilot project.
Existing Digitized Assets

A closer look into ApplicationXtender and the existing fee-based imaging service revealed troubling news. The images in the database were in JPEG format, ranging from 72-dpi photographs (of the photographs) taken by a hand-held digital camera, to 300-dpi scans with many defects. The accompanying descriptive metadata include photographer, year, location, an in-house subject, and a few lines of description. The search engine is symptomatic of search interfaces of its day. For example, the description field is broken into four separate 128-character text fields, and searches could only be performed on one field at a time. (The workaround, obviously, is to perform the same search four times.) Nevertheless, the 70,000 item-level descriptions were the strength of this system. The majority of patrons had learned to work with this interface, and they generally had been able to fulfill their research needs. The system had been effective enough to keep the fee-based imaging service going.

This imaging service received roughly 400 requests annually. When a request came in, one of the librarians would retrieve the print or negative for the requested photographs, and then the technician would scan at 400 dpi and turn over TIFF images on a CD. However, ApplicationXtender cannot handle the TIFF format, and, as a result, these images were simply being deleted.

This situation was similar to the “serious image management problem” described in Novara (2010), where new digital files were continuously being scanned, and the database was inaccessible and user-unfriendly. At the same time, long-term preservation strategies and metadata specifications were being developed while waiting for a new digital contents management system to be brought online in the near future.

We took several steps similar to those described in Novara (2010): setting imaging specifications and developing nationally compliant metadata policy. We also noted the difficulties Novara (2010) encountered with the lack of hierarchical and lateral relationship between items and collections, and focused on the relevant analysis of the structure of VRA Core schema in Fox (2008). We further took into account of the significance of OAI-PMH compliance described in Colati and Colati (2011b). We also examined the dual TIFF dark archive and JPEG access derivatives arrangement in Schmidt, et al. (2011). Finally, we studied the significance of collaborative outreach in Schmidt, et al. (2011) and did some of our own.

Preparing for the Pilot Project

Imaging requirements

We adopted imaging requirements directly from the best practices document published by the now-defunct Bibliographic Center for Research Collaborative Digital Program (CDP, 2008), with some additional details from American Memory’s Technical Information site (American Memory, n.d.), most of whose guidelines have since been incorporated into FADGI (2010), and from the Arts and Humanities Data Service Guides to Good Practice site (Dunning, 2008). These imaging requirements include technical specifications such as work environment, scanner performance, calibration, color management, image processing, as well as specific instructions for quality management.

We recognized that the decision with the widest impact would be the choice of file formats.¹ For the pilot project, as well as all subsequent imaging, we chose TIFF for the digital master and JPEG 2000 for the access derivative.

TIFF is the appropriate choice for several reasons. It is a widely-accepted, open standard, non-proprietary format, which ensures interoperability and should not present complications with the anticipated digital contents management software. The format also allows for a large number of device-independent color spaces, and

¹. A detailed comparison of many image formats can be found on page 67-68 in FADGI (2010).
supports embedded technical metadata, which is a potential time-saver when scanners are configured to supply the information automatically.

JPEG 2000 is gaining increasing acceptance among software applications. In addition to data compression, which is essential for online delivery, it stands out because it can embed metadata and support multiple resolutions. The latter feature can streamline file management: one single file can support displays at multiple resolutions, from thumbnail to high-resolution.

**Selecting a Metadata Scheme**

The digital content management software would be responsible for storing the digital objects and their metadata in XML, as well as generating its own preservation and administrative metadata. So, we focused on the descriptive metadata creation process. The only material format in the Archives without an established descriptive metadata scheme was the photograph (everything else was in EAD or MARC, which could later be mapped to MARCXML or other schemas). We wanted to select an existing widely-accepted schema that could handle hierarchical relationships. This eliminated Dublin Core, and left MODS, EAD, and VRA Core 4.0. MODS, like MARC/MARCXML, supports relationships between records in its `relatedItem` element (MARC 76X-78X linking entries). This functionality is intended to establish bibliographic relationships between separately-cataloged items. For cataloging collections of photographs, these fields can take on a new role in describing item-collection, item-item, and negative-positive-print relationships. In contrast, EAD and VRA Core 4.0 support these relationships natively. However, EAD is too robust and complex compared to VRA Core 4.0’s three hierarchies: collection, work, and item. VRA Core 4.0 also provides fuller support for other art-related fields. For these reasons, we selected VRA Core 4.0 as the metadata scheme for photographs.

**Internal Outreach**

Before the pilot project began, we reached out to the IT staff and the web development team. At QBPL, there is a strong tradition of each library division maintaining its own data on network shared drives, but digital images can quickly fill up available quota, and the frequent transfer of large files can affect network traffic performance. Furthermore, storage solutions for long-term preservation of a digital repository require significant investment and active management. In our case, this infrastructure would be implemented over time, but it is never too early to notify the IT staff. In fact, the head of IT immediately instituted a more stringent backup schedule, created extra dedicated storage capacity, and ordered additional network capacity to the imaging lab in anticipation of the new imaging devices.

The web development team would be responsible for creating the public site for search, browse, and display. After coming to understand the structure of VRA Core, the web team would develop an interface with faceted search capabilities. However, the team would not develop an interim interface before the content management software was fully installed and tested.

**Implementing the Pilot Project**

**Inventory and Legacy Data**

The goal of the pilot project was to digitize and catalog 176 photographs taken by Hal B. Fullerton between 1880 and 1910, mostly on the topic of transportation—locomotives, railroads, train stations, and automobiles—in Kings, Queens, and Suffolk. This collection was chosen for several reasons—it was relatively small, the provenance was known, it was in the public domain (all photographs had been taken before 1923), the subject matter was narrow, and the collection consisted of a mix of media (prints, glass negatives, nitrate negatives, and interpositives) and sizes. The imaging technician had just scanned this collection, so

2. *Data storage recommendations can be found in Section VII of FADGI (2010).*
both high- and low-resolution images were available. With the help of systems librarians, ApplicationXtender records for the collection were exported to a spreadsheet.

The legacy data presented quite a few challenges. Accession numbers, called Control Numbers, could not be reliably sorted, for example: HBF-93, HBF-846, HBF-1409B, and HBF-5376A-2. Some items were missing the date, which is a required element in VRA Core 4.0. All other text fields, including location, photographer, subject, and description, were all in capital letters. Upon further inspection, there were quite a number of misspellings and errors in these fields. (See Table 1 above.)

The capitalization of the location, photographer, and subject fields would only need a quick fix. However, the description fields were a different matter, even after concatenating the four fields. With capitalization problems, misspellings, and inaccuracies, it became clear that not all data fields could be directly migrated, and each photograph would need to be examined and cataloged individually. The next step, then, was to establish a cataloging policy and a workflow that maximized quality, accuracy and efficiency.

Cataloging and Metadata Policy

Our cataloging policy was derived from the Anglo-American Cataloguing Rules, second edition (AACR2). The minimum level of cataloging corresponded closely with the first level of description set forth in AACR2 1.0D1: title proper, first statement of responsibility, edition, material, publisher, date of publication, extent, notes, standard number, and terms of availability (Gorman & Winkler, 1998). A simple crosswalk was devel-

<table>
<thead>
<tr>
<th>CONTROL</th>
<th>SITE</th>
<th>DATE</th>
<th>SUBJECT</th>
<th>PHOTOGRAPHER</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBF-9920D</td>
<td>Queens Village (New York, N.Y.)</td>
<td>1898</td>
<td>Long Island Railroad Company / Accidents</td>
<td>Fullerton, Hal B.</td>
<td>clearing the SITE and REMOVING the WRECKAGE at Queens Village (New York, N.Y.), AFTER DERAILMENT of the Long Island Railroad Company TRIPLE-HEADER RUSSELL WEDGE PLOW and SUBSEQUENT FIRE / NOV. 28, 1898</td>
</tr>
</tbody>
</table>

Table 1. Excerpts from data exported from ApplicationXtender
oped between these AACR2 elements and VRA Core 4.0 (as well as to Dublin Core for future work on OAI-PMH compliance) as shown in Table 2 above. Elements applicable to photographs in terms of VRA Core 4.0 included title, agent, material, date, measurements, and description/inscription/location, and attributes refid and href. These were the metadata we would collect in the cataloging process.

One aspect AACR2 does adequately not cover is hierarchical relationships. VRA Core 4.0 is designed to distinguish and relate three hierarchical levels: collection (collections of art works), work (the actual art works), and images (visual reproductions of art works). The VRA 4.0 definition of work and image present some challenges. As defined in the schema’s documentation, work is "a unique entity such as an object or event," and image is "a visual representation of a work in either whole or part" (VRA...Introduction, 2007). In the element description, work is further refined as "a built or created object," and image "a visual surrogate of such objects" (VRA...Element Description, 2007). However, the distinction between work and image as applied to photography is not as clear. This topic sparked long discussions with the library’s cataloging staff. Is the photographic print a work? What about the negative? If the print is a work and the negative is an intermediary, is the negative, then, an image? But the negative is obviously not an image of the print. So, is it a related work? Can the negative and the print be separate works? The final verdict was to consider the work as a “created object” in the abstract sense: the photographer’s vision of the capture of a particular moment (or moments) in a particular field (or fields) of vision. (The plurals were to include multiple exposures, continuous exposures, and stereoscopic photography.) Construed in this way, tangible forms of these visual captures, such as negatives and prints, were all considered images. This demarcation worked particularly well for the Archives’ photograph collections because they were mainly documentary—photographs of buildings, railroad stations, locomotives, street scenes, etc.; the negatives and prints were not artistically produced—many were commercially manufactured and developed.

These definitions for work and image also worked well for catalogers, because catalogers could draw parallels to the cataloging concepts of manifestation and expression. This division also conveniently separated all the content description and subject analysis in work records from physical characteristics in image records. This separation would later influence workflow significantly.

**Controlled Vocabulary**

Other than the free-text photograph description and the notes fields, all other fields in VRA Core 4.0 descriptive metadata can be assigned to a controlled vocabulary. (The schema even provides the option for formatted dates.) In consultation with the cataloging department, we chose the Getty Research Institute’s Art & Architecture Thesaurus (AAT) as the main vocabulary, given its appropriate scope and depth, and its ability to integrate with VRA Core. (For example, URI expressions were available for external data linking). In addition, Library of Congress Authorities would be used for elements that had direct bibliographic counterparts. These elements included names (personal and corporate), and subject headings (topical and geographic). Terms that did not have an authorized form would have one created.

A controlled vocabulary for geographic location presented more complexity. Most of the Archives’ photographs were documented with street addresses. The Census Bureau’s Master Address File (MAF) would be a great choice for a controlled vocabulary, especially given the potential to convert the MAF’s address identifiers to uniform resource identifiers (URIs), which would enhance automated linking with any internet-based discovery system. However, neither the MAF nor the United States Postal Service database is available in list form. Using commercial products based on these databases or writing custom database interfaces was, unfortunately, beyond the scope of our program. Also, there were issues of general areas identified only by landmarks, obsolete street names, obsolete numbering, as well as non-addresses: for example, “Holtsville Station, near 985 Waverly Avenue, Holtsville, NY 11742,” where the train station was razed in 1962; “Buhrman’s Store, Bayside, near Alley Pond, now near West Alley Road and East Hampton Boulevard, Oakland Garden, NY 11362,” where neither the store, the pond, nor the road where the store once stood exists today.
Even though we could not resolve the address problem when we were working on the pilot project, we recorded full addresses with structured punctuation and keywords such as “and” (to denote intersection), “near,” and “now.” This information would be sufficiently formatted for machines to parse, should a controlled vocabulary be instituted in the future.

With geographic data processing in mind, we decided to include longitude and latitude information as well. These geographic coordinates could be obtained through an online tool such as Google Maps, which catalogers would most likely be using anyway to verify addresses during the cataloging process. VRA Core 4.0 would be able to accommodate these data through the extent attribute.

We made several other decisions about controlled vocabulary: using inches for dimensions, since the majority of the photographic papers were 3”x5”, 4”x6” or 8”x10”, and conversion to centimeters could be automated; setting the level of granularity to distinguishing between black-and-white and color prints and between negatives from interpositives; and using “digital” as the material type for scanned images.

**Data Dictionary**

A data dictionary is an essential reference for ensuring uniform use of schema elements, especially when cataloging responsibilities will be distributed to a team of librarians. The one we developed was derived from the VRA Core 4.0 data dictionary (VRA…Element Descriptions, 2007) with additional information specific to DAMS. Controlled vocabularies were specified for their respective elements; frequently-used elements and attributes were highlighted; a commentary area was added to explain some of the rationales and intended scope for each element, sub-element, and attribute. We hoped that all this information would ensure uniformity and enable sound judgments.

Based on this data dictionary, a full crosswalk from VRA Core 4.0 to simple Dublin Core was also developed for future OAI-PMH compliance. DAMS was among the first to develop this crosswalk, but it is now included in VRA Core 4.0 Element Description (2007). An excerpt is shown in Table 3 on the next page.

**Workflow and Automation**

As we created descriptive metadata for the pilot project, the workflow and data entry interface for metadata creation were developed along the way. First, a number of global regular expression search and replace routines were built to migrate the legacy data from ApplicationXtender to useful form. They were applied to the Control Number, photographer, date, and subject headings in order to correct capitalization and formatting. Similar routines were applied to output names and subject headings in LC-authorized form. This method was able to correct the majority of the entries, which would save catalogers from having to correct each entry manually. We tested the method during the pilot project. In the future, these routines can be re-applied to all other legacy data.

The photograph description fields contained so many inaccuracies and errors that they were not reusable. However, these descriptions had originally been supplied by expert historians who identified the time, location, event, people, and objects in many photographs. Therefore, we decided to transfer the legacy description field directly to the new records but only make it visible to catalogers. Catalogers, then, would create a new description and supply a unique title.

There were two physical characteristics of each photo that we recorded. First, the material of the negatives and the prints; second, the dimensions. These data were recorded in the image records.

When we looked at the cataloging process as a whole, we found a distinct difference between the expertise for creating a work record and an image record. In the work record, we needed to inspect name and subject headings for the LC-authorized form, to create a free text description, and to supply a title. This would be a process
Table 3. Excerpts from the DAMS Photographs Data Dictionary.

<table>
<thead>
<tr>
<th>Section</th>
<th>Wrapper Name</th>
<th>XML Element Name</th>
<th>XML Sub-element Name</th>
<th>XML Attribute Name</th>
<th>Description</th>
<th>Data Type</th>
<th>Authority (local decisions)</th>
<th>Corresponing Dublin Core Element</th>
<th>Mandatory?</th>
<th>Unique?</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td></td>
<td></td>
<td></td>
<td>Contains elements that describe the names, appellations, or other identifiers assigned to an individual, group, or corporate body that has contributed to the design, creation, production, manufacture, or alteration of the work or image.</td>
<td>none</td>
<td>container</td>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>display</td>
<td></td>
<td></td>
<td>A free text note about AGENT.</td>
<td>Free text</td>
<td>none</td>
<td>If there is more than one agent, the attribute extent is used.</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td></td>
<td></td>
<td>Contains elements that describe a single agent.</td>
<td>Element</td>
<td>none</td>
<td>CREATOR; CONTRIBUTOR</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>extent</td>
<td></td>
<td>Qualification of the &lt;role&gt; subelement.</td>
<td>Free text</td>
<td>LCSH</td>
<td>Describes the part of the works or images that the agent is associated with.</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>vocab</td>
<td></td>
<td>Describes the controlled vocabulary source from which extent is recorded.</td>
<td>Free text</td>
<td>none</td>
<td>Example: LCSH</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Table 3 Part 1 continued

<table>
<thead>
<tr>
<th>AGENT</th>
<th>agentSet</th>
<th>agent</th>
<th>--</th>
<th>vocab</th>
<th>Describes the controlled vocabulary source from which extent is recorded.</th>
<th>Free text</th>
<th>none</th>
<th>Example: LCSH</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>name</td>
<td>--</td>
<td>Name, appellation, or other identifier assigned to an individual, group, or corporate body that has contributed to the design, creation, production, manufacture, or alteration of the work or image.</td>
<td>Free text</td>
<td>LC or LC style</td>
<td>--</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>name</td>
<td>type</td>
<td>Qualification of &lt;name&gt;.</td>
<td>Free text</td>
<td>VRA 4.0 Restricted schema</td>
<td>Data values: “personal”; “corporate”; “family”; “other”</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>culture</td>
<td>--</td>
<td>Name of the culture, people (ethnonym), or adjectival form of a country name from which the Collection, Work or Image originates, or the cultural context with which the Collection, Work, or Image is associated.</td>
<td>Free text</td>
<td>none</td>
<td>Currently unused. Use only for describing the AGENT. Use CULTURAL CONTEXT when describing a Collection, Work, or Image.</td>
<td>COVERAGE</td>
<td>No</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>dates</td>
<td>--</td>
<td>Contains elements for date or range of dates associated with &lt;name&gt;.</td>
<td>Element container</td>
<td>none</td>
<td>Use only when referring to the Agent. Use DATE; COVERAGE</td>
<td>DATE;</td>
<td>No</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>dates/earliestDate</td>
<td>--</td>
<td>Date of birth of an individual or head of family; date of founding of a corporation.</td>
<td>Free text</td>
<td>none</td>
<td>Format: YYYY, YYYY-MM, or YYYY-MM-DD.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>dates/latestDate</td>
<td>--</td>
<td>Date of death of an individual or the last member of a family; date of closing of a corporation.</td>
<td>Free text</td>
<td>none</td>
<td>Format: YYYY, YYYY-MM, YYYY-MM-DD, or present.</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>dates/earliestDate or latestDate</td>
<td>--</td>
<td>Use for approximate dates.</td>
<td>Free text</td>
<td>VRA 4.0 Restricted schema</td>
<td>Data values: “true”; “false”</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>AGENT</td>
<td>agentSet</td>
<td>agent</td>
<td>role</td>
<td>--</td>
<td>Qualification of &lt;name&gt;.</td>
<td>Free text</td>
<td>AAT</td>
<td>Use the singular form.</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
familiar to catalogers. For the image record, the tasks of selecting the appropriate AAT material term and supplying dimensions (or dpi for the digital surrogate) could be performed by paraprofessionals. A time-saving strategy would be to have the imaging technician perform those two tasks at the time of scanning. In fact, the cataloger’s and the technician’s parts need not take place in any particular order. With the help of the old descriptions, catalogers could even proceed with the digital surrogate alone. They would need to examine the actual photographs in only a small number of cases. This strategy would work as long as the work and image records could be merged. The possibility of catalogers and technicians working in parallel opened up many possibilities for the workflow, and influenced the automation and data input strategies.

**Record Identifiers and File Naming**

File names can serve as convenient identifiers for processing and retrieval. However, they cannot serve as long-term metadata records, because their construction is limited by the file system, and files can be renamed. The format of the file names should also strike a balance between being useful and being too long and complex. In the case of the Archives’ photographs, we wanted file names to reflect each photograph’s Control Number, which is a unique accession number. Many of these numbers, however, needed leading zeros added so that they would be sortable. Since none of the photograph collections had more than 10,000 items, four digits were used. We added the leading zeros using a regular expression search and replace directly on the legacy data and replaced all file names using a freeware utility called the Bulk Rename Utility. Then we gave all the files a uniform three-letter extension that reflected the file format—.tif and .jpg. An added benefit to this file naming convention is that it facilitates any future automatic processing—Control Numbers can be extracted from file names, and file names can be constructed by extracting the Control Number from the record.

The VRA Core 4.0 schema requires an XML identifier for each record. In our case, conveniently, the newly formatted Control Numbers could be used as the record ID for work records, and the file name could be used in image records.

**Data Entry**

One other question that arose during the manual cataloging process was the data entry method. The metadata librarian began exclusively in an XML editor, but it would not be cost-effective to train every cataloging staff in the principles of XML and working directly with VRA Core 4.0 in its native XML form. The solution was to create data entry forms. After a few trials, neither Microsoft Word’s field coding nor Excel’s text export function resulted in a satisfactory data entry environment or output. Microsoft Word supports XML natively, and can handle some conditional situations (such as handling a variable number of subject fields), but it only checks for well-formedness, not for validity of the document. XML tags cannot be protected in Word, and it is quite cumbersome to work with tags if they are accidentally altered or erased. As for Excel, the usual method is to set up a spreadsheet with appropriate input fields while protecting the tags in surrounding fields, and then to export the spreadsheet as a text file. However, Excel’s text export adds spaces between cells and unwanted quotation marks around text fields.

In our context, although these problems were not insurmountable, we had another goal, which was to maximize efficiency. VRA Core 4.0 presented a challenge, because the schema was designed to hold machine-readable data as well as human-readable expressions of the same data. For example, for a 5-inch by 7-inch black-and-white print, the numbers 5 and 7, and the AAT term “black-and-white print (photograph)” would appear in the measurements and material element sets. Then the display tags would read “5 x 7 inches” and “black-and-white print” as shown in Figure 1 below. Similar situations occurred with the date, name, and description fields. Entering display tags was redundant, and would cost a considerable amount of time and increase errors. Ideally, the display tags should be automatically generated based on the data. However, this kind of text generation is not possible in Word or in Excel.

Figure 1. Sample descriptive metadata and their expression in VRA Core 4.0

<table>
<thead>
<tr>
<th>Control Number</th>
<th>How many?</th>
<th>Width</th>
<th>Height</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBF-0092</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>black-and-white negative</td>
</tr>
<tr>
<td>HBF-0092</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>interpositive</td>
</tr>
<tr>
<td>HBF-0092</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>black-and-white print (photograph)</td>
</tr>
</tbody>
</table>

With that in mind, we turned our focus to developing text processing scripts while finding the most efficient interfaces for collecting only the data. For our imaging technician, the solution was straightforward. The inputs were width, height, and material. We set up a spreadsheet with these three fields, and the technician filled them out while scanning the item. We named it the Inventory Form, which was essentially the spreadsheet shown in Figure 1 above. For the catalogers, a similar spreadsheet proved to be too wide, due to the number of free text fields, and was difficult to see and navigate on the screen. We found another solution in Microsoft Access. We populated the same spreadsheet with an Access Form object that had a much cleaner and more intuitive interface. We named this interface the Descriptive Metadata Entry Form (Figure 2 on the next page) and the underlying spreadsheet the Descriptive Metadata Table. With this set up, controlled vocabularies could be directly linked to authorized fields via drop-down menus. This sped up input and reduced errors, and, it worked especially well for the pilot project, because there was only a single photographer and the collection covered a limited number of subject headings.

These two interfaces—the Inventory Form and the Descriptive Metadata Entry Form—served to collect the underlying descriptive metadata that would eventually be assembled into VRA Core records.

Data Processing and Metadata Creation

Descriptive metadata were assembled in two steps. The Descriptive Metadata Table was processed first to create valid VRA Core 4.0 records. These provisional records already had all necessary identifiers, internal relationships, and external links, thanks to the file naming convention, and could be ready for immediate public use. Then we inserted data from the Inventory Form to complete these records.

Populating metadata records with data can be done in many ways. We opted for Microsoft Word’s mail merge function for processing the Descriptive Metadata Table, because the librarians had enough familiarity with Word that they could potentially run the process on their own in the future. During the mail merge, data were extracted from the Descriptive Metadata Table and inserted between opening and closing tags. Conditional formulas made up most of the merge codes, because whether new tags would be created depended on whether
the data field was blank. A few attributes were also dependent on the data. For example, *circa* was evaluated based on the “yes” or “no” value in the “circa” column. An excerpt of this merge document is shown in Figure 3.

Figure 3. Excerpts from the Microsoft Word merge document that assembled data from the Descriptive Metadata Table and created provisional VRA Core 4.0 records.
For the 176 records of the Hal B. Fullerton Photographs, this merge process took about 5 seconds. Once these provisional records were created, we could then insert inventory data as they became available. The process of traversing and inserting XML tags was beyond the capability of mail merge, and a text processing script was the only option. We chose Python as the language, because it has built-in functions for hierarchical data structures like XML. The text processing script read the Inventory Form and inserted data into the appropriate tags. The script also examined the numeric data and controlled vocabulary, and composed the free text descriptions in the display elements for onscreen presentation. Returning to Figure 1, the script first inserted the data: width ("7," "5," and "10"), height ("5," "4," and "8"), and dpi ("400") in measurements, and the AAT terms ("black-and-white negative," "interpositive," "black-and-white print (photograph)," and "digital image") in material. Then, based on each value in material, the script filled the refid attribute with AAT’s unique numeric identifier, and linked measurements to material through the extent attribute. Then, based on the three values in the “How many?” column, measurements, and materials, the script composed the phrases in the two display tags. Furthermore, the script also inserted a depictedIn relation in the work record to show the materials represented in the image record. Figure 5 on page 45 shows two examples of full records created by this process; the full workflow diagram is shown in Figure 4 on the next page.

Additional Considerations

Finally, we wanted to ensure the sustainability of the metadata, and to be prepared for worst-case scenarios. What if the content management software company were to go out of business? What if we were to lose funding for technical support and the software malfunctioned? We wanted to have the capability to set up a website based on existing metadata and files alone without the content management software. This had already been accomplished by basing file names and basing XML identifiers on Control Numbers, and having external links automatically generated in the metadata creation process. These records and files were designed to contain sufficient information such that all search, retrieval, and online presentation functions could be performed by drawing information from the XML records alone, processing script such as Python scripts or XSLT (Extensible Stylesheet Language Transformations).

We performed two experiments to test search and retrieval operations on the Hal B. Fullerton records. In the first experiment, we ran our records through the XSLT stylesheet used for showing VRA examples on the Visual Resources Associations web site. The browser displayed the resulting HTML as expected. The second experiment simulated a search and filter by location. We wrote a Python script to parse the addresses and geographic coordinates from the metadata records, and then search for sets of photographs that were taken at the same location and group them together. To visualize this, we used the Python script to export the results to an XML file in the Keyhole Markup Language (KML), a schema used for displaying data in an earth browser such as Google Maps. The script processing took about 15 seconds, and the resulting display is shown in Figure 6 on page 47.

Conclusion

This pilot project served successfully as a testing ground and enabled us to develop cataloging policy documents, a metadata creation workflow, and useful tools and scripts. The experience will serve as the foundation for metadata creation at the production scale, and as the basis for future changes and improvements to the digitization process.
Figure 4. Flow chart showing metadata creation workflow for archival photographs.
Figure 5. Two sample VRA Core 4.0 records.

<work id="HBF-9920D" source="Hal B. Fullerton Photographs">
  <dateSet>
    <display>1898</display>
    <notes>-------- This is record no. HBF-9920D --------</notes>
    <date type="creation">
      <earliestDate>1898</earliestDate>
      <latestDate>1898</latestDate>
    </date>
  </dateSet>
  <descriptionSet>
    <display>Removal of wreckage following derailment of Long Island Rail Road Company a 4-4-0 steam locomotive that was pushing a triple-header Russell wedge plow at Queens Village, November 28, 1898.</display>
    <description source="Xtender">clearing the site and removing the wreckage at queens village (new york, n.y.), after derailment of the Long Island Rail Road Company triple-header russell wedge plow and subsequent fire / nov. 28, 1898 / print purchased from ron ziel.</description>
    <description source="KS">Removal of wreckage following derailment of Long Island Rail Road Company a 4-4-0 steam locomotive that was pushing a triple-header Russell wedge plow at Queens Village, November 28, 1898.</description>
  </descriptionSet>
  <locationSet>
    <location type="creation">
      <name type="geographic">Queens Village, NY 11429 (40.717703,-73.73597)</name>
    </location>
  </locationSet>
  <relationSet>
    <relation href="HBF_collection.xml" refid="HBF-9920D" type="partOf">Hal B. Fullerton Photographs</relation>
    <relation relids="HBF-9920D.tif" href="URI of Image" refid="HBF-9920D" type="depictedIn">negative; interpositive; print; digital image</relation>
  </relationSet>
  <rightsSet>
    <display>Public domain</display>
    <rights type="publicDomain"/>
  </rightsSet>
  <subjectSet>
    <subject vocab="LCSH">
      <term type="descriptiveTopic" vocab="LCSH">Railroad snowplows</term>
      <term type="descriptiveTopic" vocab="LCSH">Steam locomotives</term>
      <term type="conceptTopic" vocab="LCSH">Railroad accidents</term>
      <term type="geographicPlace" vocab="LCSH">Queens Village (New York, N.Y.)</term>
      <term type="corporateName" vocab="LCSH">Long Island Railroad Company</term>
    </subject>
  </subjectSet>
  <titleSet>
    <title source="KS">Wreckage of a Snow Plow Train</title>
  </titleSet>
</work>
<work id="HBF-9970" source="Hal B. Fullerton Photographs">
  <dateSet>
    <display>ca. 1900</display>
    <notes>------- This is record no. HBF-9970 -------</notes>
    <date type="creation">
      <earliestDate circa="true">1900</earliestDate>
      <latestDate circa="true">1900</latestDate>
    </date>
  </dateSet>
  <descriptionSet>
    <display>LIRR ferry terminal, coal chute and power station from the East River.</display>
    <description source="KS">LIRR ferry terminal, coal chute and power station from the East River.</description>
  </descriptionSet>
  <locationSet>
    <location type="creation">
      <name type="geographic">Long Island Rail Road Ferry Terminal, West 2nd Street and East River, now near Borden Avenue and 2nd Street, Long Island City, NY 11101 (40.741822,-73.961307)</name>
    </location>
  </locationSet>
  <relationSet>
    <relation relids="F012" href="HBF_collection.xml" refid="HBF-9970" type="partOf">Hal B. Fullerton Photographs</relation>
    <relation relids="HBF-9970.tif" href="URI of Image" refid="HBF-9970" type="depictedIn">negative; digital image</relation>
  </relationSet>
  <rightsSet>
    <display>Public domain</display>
    <rights type="publicDomain"/>
  </rightsSet>
  <subjectSet>
    <subject vocab="LCSH">
      <term type="descriptiveTopic" vocab="LCSH">Power-plants</term>
      <term type="descriptiveTopic" vocab="LCSH">Waterfronts</term>
      <term type="conceptTopic" vocab="LCSH">East River (N.Y.)</term>
      <term type="geographicPlace" vocab="LCSH">Long Island City (New York, N.Y.)</term>
      <term type="corporateName" vocab="LCSH">Long Island Railroad Company</term>
    </subject>
  </subjectSet>
  <titleSet>
    <title source="KS">Coal Chute and Power Station</title>
  </titleSet>
</work>
Figure 6: KML output from the Python script processing of the VRA Core 4.0 metadata record displayed in Google Maps, showing the geographic distribution of the Hal B. Fullerton Photographs: blue pegs represent locations where photographs were taken; pop-up balloons display thumbnail images and descriptive metadata.
tization program. We look forward to working with content management software and with new imaging equipment, as well as future development of the web interface and further integration with geographic information systems.

Bibliography


DIGITIZING EVERYTHING: PILOTING METADATA CREATION
