

## **"Where the Giants Stand: Protecting the Public Domain in Digitization Contracts with Commercial Partners"**

*Richard Fyffe, Assistant Dean for Scholarly Communications, University of Kansas Libraries*

*Beth Forrest Warner, Director, Digital Library Initiatives, University of Kansas*

### **Abstract:**

Digital dissemination of public-domain works of historic, literary, and artistic value increases the intellectual value of those materials; repositories that license digitization rights to commercial agencies should seek the greatest possible availability for these works on the open Internet. As with copyright law, a balance must be struck between society's legitimate interest in maximizing access to and use of the work and society's equally legitimate interest in encouraging capital investment in digitization, dissemination, and long-term sustainability. This paper proposes that the open-access movement and efforts to "reclaim the public domain" provide a theoretical framework for evaluating prospective partnerships between non-profit repositories and commercial digital publishers, and recommends practical guidelines for developing digitization contracts that both uphold the value of the public domain and meet the needs of the marketplace.

### **Keywords (3-10):**

Public domain  
Digitization  
Contracts  
Licenses  
Open access  
Commercial vendors

### **Author Information:**

Richard Fyffe, Assistant Dean for Scholarly Communications, University of Kansas Libraries, 1425 Jayhawk Blvd., 351 Watson Library, University of Kansas, Lawrence, KS 66049; [rfyffe@ku.edu](mailto:rfyffe@ku.edu). MA (Philosophy, University of Connecticut), MLS (Simmons College), BA (Philosophy, University of Connecticut).

Beth Forrest Warner, Director, Digital Library Initiatives, University of Kansas; Information Services, 1450 Jayhawk Blvd, 223 Strong Hall, University of Kansas, Lawrence, KS 66049; [bwarn@ku.edu](mailto:bwarn@ku.edu). MLS (University of Denver), BA (Chemistry, College of Wooster)

"Where the Giants Stand: Protecting the Public Domain in Digitization Contracts with Commercial Partners"

## "Where the Giants Stand: Protecting the Public Domain in Digitization Contracts with Commercial Partners"

"If I have seen further it is by standing upon the shoulders of Giants"--Isaac Newton to Robert Hooke, 5 February 1675.

"[...] I find that principles have no real force except when one is well fed..."--Mark Twain, *Extracts from Adams's Diary*.

### *I. Introduction*

Two rapidly growing movements are reshaping scholars' expectations for the availability in digitally networked form of intellectual works created or curated with public support. One is a movement to "reclaim" the public domain, or protect it by limiting measures that seek to extend copyright protection to works that could otherwise be freely copied and modified by all users. The other is a movement to assure the universal cost-free availability of peer-reviewed scholarly articles.

In this paper, we will argue that these movements create a new context within which libraries, archives, and other repositories of intellectual and cultural heritage should evaluate their strategies for providing digital access to their holdings of literary, historical, and artistic works in the public domain. For such public-domain works, the institution's ownership of the tangible medium in which the work is instantiated gives it a limited set of property rights, but not the *exclusive* right to disseminate the intellectual property or use it to create new works.<sup>1</sup> Although many successful digital projects have been publicly or privately funded and then disseminated without direct cost to users,<sup>2</sup> partnership with a commercial publisher is naturally seen as an attractive way of covering the costs of digitization, dissemination, and marketing while extending access to these objects and safeguarding them from the damage that may result from direct use. In such partnerships, a repository licenses (for some form of compensation) to the commercial partner the right to digitize from its holdings, and the commercial partner in turn licenses access to paying subscribers, expecting a fair financial return on its investment and its risk -- and therefore restricting access to those subscribers.<sup>3</sup> The fact that subscription-based licensing is becoming a primary mode of information provision for many libraries and other repositories increasingly gives such an arrangement an appearance of business-as-usual. Such partnerships with commercial publishers may appear to be an attractive way for the repository to

gain access to capital, specialized equipment, and trained personnel, and to share (or even eliminate) the financial risk of digitizing materials that may turn out not to have a large audience.

The open-access and public-domain movements, we will argue, call into question (we do *not* argue that they invalidate) the propriety of library/commercial partnerships in creating and disseminating digital collections of public-domain materials. Our paper has two parts. In the first, more theoretical, part (section II) we will briefly review the arguments associated with both of these movements and note the ways in which they do and do not apply to the case of libraries and archives that hold public-domain materials that they wish to digitize and disseminate. Throughout this paper we will use the term "repository" to cover the various kinds of institutions that may hold materials to which the concept "public domain" applies, including libraries, historical societies, and archives. It should be noted that although the materials targeted by academic libraries for digitization are often held in their rare book or special collections units, we will not be concerned with the special administrative issues that may apply there.<sup>4</sup> Public-domain material of great intellectual value may also be found in the general or circulating collections of most libraries, and the arguments we advance here apply to both closed-stack and open-stack collections.

In making these comparisons, we will acknowledge the economic constraints facing repositories, the costs associated with successful and sustainable digital projects, and the legitimate role that may be played by commercial firms seeking profit in return for capital investment in these projects. We will also recognize that subscription-restricted access to digital copies of public-domain works does not change the copyright status of these works; the works themselves and the originals of which the digital objects are copies remain available to the public for copying, publishing, and using them to create derivative works. Nevertheless, we will argue for the growing importance of the world-wide digital commons created by the Internet, and suggest that the "opportunity cost" of restricting access to digitized versions of the materials of intellectual heritage must be factored into a decision to use a subscription model to defray the costs of conversion and dissemination. In this respect, we will argue, digital dissemination differs essentially from traditional forms of publication for reformatted materials like microform publication.

In the second part of the paper (section III), we will apply this perspective by suggesting some considerations that should guide libraries and other repositories in a decision to license digitization rights to a commercial partner (these considerations may apply as well to a repository that chooses to use subscription fees to defray the internal costs of digitization and dissemination); and we will recommend terms that should be incorporated into a digitization license.

## *II. Toward the Digital Commons: The Public Domain and Open-Access Scholarship*

Article I, Section 8 of the United States Constitution empowers Congress to "promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." This language is traditionally interpreted as calling for a balance between the rights of authors and inventors to be fairly compensated for their intellectual labors and the rights of society at large to use the products of those labors for the creation of new knowledge and new technology. For a prescribed period of time, authors and inventors hold the near-exclusive right to disseminate their works and create derivative works, but after that period has elapsed those works may be freely disseminated and used by anyone (they enter the public domain). Moreover, even during that period of copyright protection, society at large enjoys certain limited rights to use the work, under conditions defined as "fair use".<sup>5</sup>

Statutory and case law define the period during which a work may enjoy copyright protection and the kinds of uses of copyright-protected work that will be considered "fair." A growing number of legal scholars, artists, and activists are expressing concern over the effect of a long series of extensions to the period of copyright protection. The duration of copyright protection was originally set at 14 years, with the possibility of renewal for another 14 years if the author was living. The period was extended by legislation in 1831 and 1909. In 1976, the Copyright Revision Act further extended protection (for works created after January 1, 1978) to 50 years after the death of the author. In 1998 Congress passed the Sonny Bono Copyright Term Extension Act (CTEA), extending the term of all existing copyrights by an additional 20 years.

CTEA thus extended the term of most existing copyrights to 95 years and that of many new copyrights to 70 years after the author's death.

In a suit filed in 1999, Eric Eldred challenged the constitutionality of CTEA. Eldred argued that the effect of this series of copyright extensions has been to render copyright protection "perpetual," thus contravening Article I, section 8, of the Constitution. The United States Supreme Court, in an opinion delivered on January 15, 2003, ruled against Eldred, finding that "Congress acted within its authority and did not transgress constitutional limitations."<sup>6</sup> However, Justices Breyer and Stevens separately delivered sharply worded dissents from this verdict. In Justice Breyer's words, "The economic effect of this 20-year extension--the longest blanket extension since the Nation's founding--is to make the copyright term not limited, but virtually perpetual. Its primary legal effect is to grant the extended term not to authors but to their heirs, estates, or corporate successors. And most importantly, its practical effect is not to promote, but to inhibit, the progress of 'Science'--by which word the Framers meant learning or knowledge."<sup>7</sup>

Despite the failure of *Eldred v Ashcroft* to void the Copyright Term Extension Act, a growing movement of judges, legal scholars, artists, and activists is drawing national attention to the importance of public domain works as the foundation for the creation of new knowledge and creative work. A petition campaign, Reclaim the Public Domain, argues that serial extensions of copyright protection "unnecessarily threaten the public domain without any corresponding benefit to copyright holders" and calls on Congress to enact new legislation, the Public Domain Enhancement Act, that would

require American copyright owners to pay a very low fee (for example, \$1) fifty years after a copyrighted work was published. If the owner pays the fee, the copyright will continue for whatever duration Congress sets. But if the copyright is not worth even \$1 to the owner, then we believe the work should pass into the public domain. This legislation would strengthen the public domain without burdening copyright owners. It would also help clarify rights over copyrighted material, which in turn would enable reuse of that material. The law could thus help restore balance to the protection of copyright, and support the public domain.<sup>8</sup>

Similarly, Stanford University law professor Lawrence Lessig has argued in his best-selling book *The Future of Ideas* that "always and everywhere, free resources have been crucial to innovation and creativity" and that "without them, creativity is crippled."<sup>9</sup> In the digital age, he continues

the central question becomes not whether government or the market should control a resource, but whether a resource should be controlled at all. Just because control is possible, it doesn't follow that it is justified. Instead, in a free society, the burden of justification should fall on him who would defend systems of control.<sup>10</sup>

In his book, Lessig analyzes the social and economic conditions that foster creativity and innovation. Economic incentives--the right and the ability to profit financially from one's innovations--are one necessary component, he says. Equally necessary, however, is the right and the ability to use the works of one's predecessors: "Free content," he notes, "is crucial to building and supporting new content."<sup>11</sup> Quoting Judge Alex Kozinski of the Ninth Circuit Court of Appeals, "Creativity is impossible without a rich public domain;" "Overprotection stifles the very creative forces it's supposed to nurture."<sup>12</sup>

Alongside the movement to protect the domain of commonly held intellectual works against restrictions on *use*, a parallel movement is focused on protecting or expanding *access* to scholarly literature. According to proponents within this movement, peer-reviewed research papers are a public good that should be made as widely available as possible. Moreover, Internet-based technologies make it possible to remove print-based barriers to access:

An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds. Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge.<sup>13</sup>

Although not (or not yet) the predominant model for disseminating scholarly literature, open access is gaining momentum. The *Lund Directory of Open-Access Journals*, for example,

indexes over 550 peer-reviewed subscription-free journals in a wide range of disciplines.<sup>14</sup> The Public Library of Science, which started as a petition drive signed by over 30,000 scientists world-wide calling on conventional subscription-based journals to open their backfiles to cost-free access, was awarded a \$9-million grant in December, 2002, to develop a competing suite of new open-access journals in the biomedical sciences.<sup>15</sup>

The arguments of the proponents of open access generally turn on three key features of this literature.<sup>16</sup> In the first place, they note, creation of much scholarly literature is funded through taxpayer support either directly or indirectly. Much of the research in question is directly funded by federal agencies such as the National Science Foundation and the National Institutes of Health. Moreover, the salaries of many of these researchers (if not covered directly by public grants) and of the researchers who provide peer-review services are paid by universities and other institutions that enjoy tax-exempt status as charitable organizations. It is wrong, these proponents argue, that the host institutions that provide the research and review services should have to pay again for access to this literature through subscription fees. It is especially problematic that commercial journal publishers should realize high profits through ownership of the intellectual property represented by these papers.

Within the open-access movement, proponents differ with respect to the proper copyright status of such publicly funded works. Signatories of the Budapest Open Access Initiative expect *authors* to retain copyright to their works and call on journals to "no longer invoke copyright to restrict access to and use of the materials they publish" but instead to "use copyright and other tools to ensure permanent open access to all the articles they publish."<sup>17</sup> Under this regime, authors will not transfer copyright to journals but instead will grant specific, limited licenses to publishers for the right to disseminate their papers. By contrast, legislation introduced by Rep. Martin Sabo (D-Minnesota) in June 2003 (the "Public Access to Science Act") and endorsed by the leadership of the Public Library of Science directs that copyright protection will not be available "for any work produced pursuant to scientific research substantially funded by the Federal Government." "It is the sense of the Congress," according to this Act, "that any Federal department or agency that enters into funding agreements [for scientific research] should make every effort to develop and support mechanisms for making the published results of the research



conducted pursuant to the agreements freely and easily available to the scientific community, the private sector, physicians, and the public."<sup>18</sup>

Second, according to the proponents of open access, scholarly literature is intended by its authors to be given freely to the journals that will publish it. The authors seek compensation from their institutions in the form of salary increases, promotions, and other recognition, but they do not seek (and do not receive) compensation from the publishers of these journals.

Finally, the goal of an author of this kind of literature is to maximize its scholarly impact (typically measured by citations from subsequent papers), and this is accomplished by assuring its widest possible availability. Access to literature that is intended to generate royalties (for example, textbooks, novels, cookbooks) must be restricted to paying customers, but to restrict access to literature intended only to be read and not to generate revenue defeats its purpose.

The public-domain literary, historical, and artistic works that are the subject of this paper share some of these features of scholarly research papers and to that extent are susceptible to the same claims for cost-free digital access, but they also differ in important respects. We will examine the similarities and the differences in turn.

The repositories that hold these works typically enjoy public support, directly or indirectly. The repositories are often tax-exempt charitable organizations. They may receive a portion of their operating revenue from public sources, and they may receive special grants from public sources (such as the National Endowment for the Humanities or the Institute for Museum and Library Services) for the acquisition or care of these materials. And, if the grants or gifts are from private foundations, those foundations may themselves be tax-exempt charitable organizations.

Moreover, the tax-exempt network within which these repositories operate enlarges the support available for these holdings at public expense: taxation of the repository or its benefactors would reduce the total amount of support available for acquiring or curating the holdings, and exemption of these institutions from taxation requires that other members of society pay a larger portion of taxes in addition to the tax-support directly supplied to the repositories. These

institutions enjoy public support because they are considered to provide a significant public benefit. It diminishes this benefit to restrict access to the holdings of these institutions, especially when a commercial business is permitted to profit through subscriptions or sales that restrict public access.

Furthermore, the works in question have passed from copyright protection into the public domain because society considers the authors to have had ample time and opportunity to realize profit from their intellectual labor. (Some works lack copyright protection even at the moment of creation. Under United States law, works created as part of federal employment are in the public domain.) For these works, the profit-making period has ended.

These considerations should cause a repository of public-domain materials to study carefully any plan to license digitization rights to a profit-making firm that will restrict access to the digital copies. Is such an arrangement consistent with the public support the repository may enjoy? Should digital access to works in the public domain, works in the custody of tax-exempt and tax-supported institutions, be restricted to paying subscribers? Is the move to profit-based partnerships contrary to the movements to maintain a rich and diverse domain of copyright-free works and to open access to recent works of original scholarship? In Section III of this paper we will suggest some practical ways to help assure that digitization partnerships respect the spirit of open access and preservation of the public domain.

However, works in the public domain differ from current works of original scholarship in some important ways, and the parallels just adduced require some important qualifications.

Unlike the majority of contemporary works of original scholarship that are the target of the open-access movement, historical works in the public domain do not enter the world in digital format. As the best-informed proponents of open-access scholarship recognize, digitization and digital dissemination are neither cost-free nor cheap. A decade of experience has demonstrated that the practices necessary to digitize a collection of tangible works and then render the collection discoverable, interoperable with related collections, and capable of being migrated forward to

new technologies require time, highly trained personnel, and specialized software and equipment.

Moreover, public support for repositories is not total support. Most repositories that hold primary-source collections (and for which licensing of rights to a commercial partner is permitted) do not receive their whole operating budget from public funds. Society expects them to generate revenue independently, and partnerships that transfer resources (financial or in-kind) to the repository should, all other things being equal, be encouraged.

Finally, public-domain status for a work does not mean that profit on that work is forbidden, only that an exclusive right to profit from the work cannot be claimed. After all, traditional re-publication of public-domain works like those of Shakespeare and Dickens is a thriving business.

Given these differences from the scholarly literature that is the focus of the open-access movement, it might therefore be argued that digital dissemination of public-domain material does not differ substantially from the traditional practice of selling microform copies either directly or through a licensing agreement with a commercial microform publisher. Under such arrangements, it might be argued, access is expanded, not diminished. In the first place, distribution of microform copies to repositories across the world lessens the financial burden on scholars who would otherwise have to travel long distances to examine rare or unique artifacts. The cost to the repositories that purchase the microfilm is small compared to the collective savings enjoyed by users. Subscription-based digital distribution accomplishes the same thing in principle, with greater practical benefits such as improved searchability, accessibility from remote locations, etc.

Second, distribution of microform copies reduces handling of the original documents, preserving them for use by those few scholars (in present and future generations) who might need access to details not captured by the reproduction. The total revenue derived from these sales could not replace the collection of originals if they were to be lost through mishandling. The same argument applies to subscription-based digital distribution. Finally, neither microform sale nor fee-based digital distribution changes the public-domain status of the individual works. They

remain available for copying and for use in the creation of derivative works, as the foundation for new knowledge and new creative works.

These arguments overlook the critical difference between analog reproduction and digital reproduction in a networked environment. The emerging digital commons represents a new phenomenon, one that raises the stakes for repositories that are considering digitization of public-domain material. "The critical feature of the Internet that sets it apart from every other network before it," Lessig writes, "is that it could be a platform upon which a whole world of activity might be built. The Internet is not a fancy cable television system; the Internet is the highway system, or the system of public roads, carrying bits rather than trucks, *but carrying them in ways no one can predict*."<sup>19</sup> Moreover, Lessig points out, "Where we have little understanding about how a resource will be used, we have more reason to keep that resource in the commons."<sup>20</sup>

The Internet has created unprecedented potential for *unpredictable* discovery and use of texts and other expressions, and this potential is enriched with each addition to the commons. The enrichment is in the form of a network effect, and it is this effect that creates a greater "opportunity cost" when digitized public-domain material is restricted to paying subscribers than is seen when the same material is distributed in microform. In economics, a "network effect" is defined as an increase in the benefit that an agent derives from a good when the number of other agents consuming the same kind of good increases. Having an email account or a telephone, for example, becomes more valuable as more people have accounts or telephones. At the same time, under conditions that display a network effect the purchase of a good by one individual indirectly benefits others who own that good -- for example by purchasing a telephone a person makes other people's telephones more useful.<sup>21</sup>

A similar analysis applies to the effect of posting open-access digitized texts, based on the web's ability to collocate texts in unexpected ways and to yield unexpected materials through keyword and other kinds of searches. The intellectual value of every text already openly available is exponentially enhanced with each new addition, and the cost of the opportunities lost when one group of public-domain works cannot be accessed alongside other groups of such works is

greater than it was under an analog regime (like microform). This cost should be factored into a repository's digitization strategy.

### *III. Building the Digital Commons: Leveraging the Power of Licenses*

We have argued that digital dissemination of public-domain works of historic, literary, and artistic value increases the intellectual value of those materials, and that repositories that license digitization rights to commercial agencies should seek the greatest possible availability for these works on the open Internet. As with copyright law, however, a balance must be struck between society's legitimate interest in maximizing access to and use of the work and society's equally legitimate interest in encouraging capital investment in digitization, dissemination, and long-term curation.

The terms of a partnership between a repository and a commercial publisher should be defined in a contract by which the repository licenses to the publisher certain rights regarding its tangible property (the books or other materials that it owns) for specific compensation, and each party sets expectations for the other. The contract is an important instrument by which a repository can help to assure that the digital commons is enriched -- both through terms that maximize access over the short and the long term and (of equal importance) through terms that give the commercial partner fair opportunity to realize a return on its investment. Thus, the terms we recommend for digitization contracts must be evaluated in the context of the specific circumstances of the repository, the works under discussion, and the publisher or agent. They must also be grounded in a clear understanding of the digitization process and the costs involved in converting, disseminating, and sustaining the resulting digital collection. And finally, projects should operate within the basic principles outlined in *A Framework of Guidance for Building Good Digital Collections* from the Institute for Museum and Library Services and the Digital Library Federation.<sup>22</sup>

#### *A. The Business Model: Assessing a Prospective Commercial Partnership*

As a first step in arriving at a balanced contractual arrangement, it is important for the repository to understand the basic business approaches and requirements of the publisher. For example, the SPARC *Declaring Independence* program for the reform of scholarly publishing calls on editors

of scholarly journals to become more knowledgeable about the finances of the publications for which they provide services. If the publisher's pricing and licensing terms prove to be impeding access to their journal, SPARC urges these editors to demand that the publisher make changes or to switch to a lower-priced publishing platform.<sup>23</sup>

Similarly, repositories that are considering a commercial digitization partnership should understand the finances behind the deal. Repositories should request from prospective commercial partners a statement of costs expected to be incurred for the project. In addition, they should consider any costs *they* may incur as well. Depending on the project, these overall costs may include the selection and preparation of originals; metadata creation and indexing; preservation and conservation of originals; production of intermediates; digitization; quality control of images and data; system / network infrastructure; on-going maintenance of digital materials; and marketing, sales, and support.<sup>24</sup>

Over the past decade, the number and variety of digitization projects has begun to provide a clearer picture of the associated costs. A recent symposium of the National Initiative for a Networked Cultural Heritage (NINCH) on cost models for digitization provides several case studies of projects done by educational institutions and a breakdown of general costs from an outsourcing vendor.<sup>25</sup> However, even with this growing body of information it is evident that digitization is still not a uniform or straightforward process and that the repository must have a clear understanding of the procedures and trade-offs involved that can affect the project's cost.<sup>26</sup> For example, a recent examination of average costs by Steven Puglia, of the US National Archives and Records Administration, showed that

on average, roughly one third of the costs are related to digital conversion, one third for cataloging and descriptive metadata, and one third for administration, quality control, etc. [ . . . ] an average cost, over three years of data, of \$29.55 per digital image (but with a range of between \$1.85 and \$96.45). Within that, itemized average costs come to \$6.50 for digitizing; \$9.25 for cataloging; and \$13.40 for administration. Adjusted for unrealistically high or low costs, the figures came to \$17.65 overall (digitizing \$6.15; cataloging \$7; and administration \$10.10).<sup>27</sup>

Given these wide-ranging cost figures, it is important to know the publisher's per-item cost estimates and justifications when evaluating any project proposal.

In addition to the direct costs of the project, the repository should also know the profit margin that is built into the proposal. Commercial partners will not engage in this work without profit, but the repository should expect to negotiate reasonable margins. Although hard-and-fast rules for profit margins are difficult to determine, some benchmarks to consider are general publishing industry profit margins, online information industry profit margins, and the Consumer Price Index.

### *B. Compensation Schedules: Planning for Independence*

The contract should set a specific term during which the publisher will have exclusive rights to market the digital files created by the publisher. The term should be sufficiently long that the publisher has a reasonable opportunity to recover its investment and generate profit. To help determine the length of any exclusive distribution term, the repository should discuss marketing plans and sales projections for the collection. At the expiration of the exclusive distribution term, the repository should have the right to disseminate the files without restriction. The publisher, too, should be able to continue selling access to the collection, but it will then be competing on the basis of its platform (search engine, presentation, and any copyrighted works it may have created as supplementary to the collection), rather than through exclusive access rights for the content.

Most commercial digitization contracts will include a royalty payment to the repository based on a percentage of the net receipts.<sup>28</sup> Repositories should keep in mind that royalties can only be generated from sales revenue and that, all other things being equal (i.e., the publisher's costs) the higher the royalty is the higher the subscription price must be -- and therefore the more restricted the access. A repository may choose to negotiate a lower royalty rate, but should get contractual assurance that the lower cost will be passed along to purchasers / subscribers.

*Part of any profit derived from the commons should be returned to the commons.* The long-term sustainability of the digital commons will require a steady stream of funding and we recommend that repositories use royalty payments to create reserves or endowments for the support of open-access digital collections. These endowments may be used to purchase and maintain systems that

will host the digital collection after the commercial contract expires, to support the migration of the digital files into new formats or onto new platforms as technology changes, or to support digitization of other public-domain works held by the repository (assuming that the sustainability of existing files is provided for). For more detailed information on sustainability models, see the section on “Sustainability: Models for Long-Term Funding” in *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*<sup>29</sup> and the CLIR report *Building and Sustaining Digital Collections: Models for Libraries and Museums*.<sup>30</sup>

### *C. Copyright, Licensing, and Subscribers’ License: Retaining Control*

Before beginning any digitization project the copyright status of the original materials should be established. In the case of the materials discussed in this article, the repository should verify that they are currently in the public domain<sup>31</sup> or specifically spell out in the contract responsibilities for obtaining copyright permission for digitization.

Once materials are in electronic form, additional copyright considerations come into play:

In order for a database to warrant copyright protection, its author must have made some original effort in the collection, selection, and arrangement of material (e.g. by indexing terms or adding keywords). This protection is irrespective of whether the individual contents of the database are original and therefore copyrightable in themselves or include factual data (where the database protection will not prevent an individual from extracting them, short of copying the selection and arrangement of the database as a whole). Currently in the U.S., if the database is unprotected by copyright law, the contents of the database may be copied unless such acts are prohibited by contract or license.<sup>32</sup>

Digitization agreements with the publisher should clearly indicate who owns the copyright to the digital files, individually and in aggregated form, as well as the rights licensed to the other party. It is important to clarify that any exclusive marketing right enjoyed by the publisher pertains to the digital files and not to the original collection that was digitized. The public-domain status of these works and the repository's tax-exempt status suggest that proposals of competing publishers should be honored.



The contract should stipulate that as soon as the digital collection is ready to market, the repository will receive (and/or be given access to) a set of the files, either for immediate use by the repository, to be held in escrow by a third party, or embargoed for a specified amount of time. The repository should also retain the right to provide access to the collection from its own server(s) to a restricted community of users, or stipulate a cost-free subscription from the publisher's server. In the case of an academic library, this community would be the faculty, students, staff and other users typically authorized under contracts for licensed resources (including walk-in users); in the case of a museum or archival repository this may be the staff of the institution and any other on-site users.

In addition, the repository should stipulate key terms for the license the publisher will establish for subscribers to the digital collection, or else negotiate substantive involvement in crafting that license when the time comes. Subscribers' licenses should follow the best practices outlined in the statements of the Association of Research Libraries,<sup>33</sup> the International Coalition of Library Consortia,<sup>34</sup> the Digital Library Federation,<sup>35</sup> and the Creative Commons.<sup>36</sup> Many libraries already have experience with licenses as subscribers; they should apply this experience to the construction of good licenses in their new role as publishing partners. Perhaps one of the most important issues to consider in this process will be determining how the responsibilities for providing archival access for subscribers will be addressed -- whether this will be the on-going responsibility of the publisher, whether the subscriber will get copies of the digital files, whether the repository will be responsible for providing access, etc.

#### *D. Getting it Done Right: Standards and Work Schedules*

We have argued that the digital commons grows more valuable as more open-access work is made available over the network. It is therefore vital that the integrity of the commons be protected over time. In addition to the establishment of endowment funds to help ensure the financial sustainability of the collection, another critical consideration is its technical sustainability. Technical sustainability is directly related to the standards and best practices followed when creating the digital files; applicable standards and best practices should be stipulated in the contract. Following good practices and open standards for digital file creation will help ensure the interoperability and migration of these files over the long run. In addition,

consideration should be given to the availability or creation of metadata (descriptive, structural, and administrative) describing the digitized files; these requirements should be outlined or referenced in the contract. Sources for information on appropriate standards and best practices include the Digital Library Federation, the Research Libraries Group, the Northeast Document Conservation Center, NINCH, and the Colorado Digitization Program.<sup>37</sup> Publications from these organizations cover topics such as acceptable file formats and image resolutions for various types of materials, text scanning (both marked-up and “dirty”), and metadata formats and standards. Repositories should ensure that the contract stipulates that copies of master files (image, audio, video), marked/unmarked text, display files, and metadata be made available by the publisher to the repository either during or at the end of the contractual period, and what media will be used.

The schedule for digitization of materials should be laid out in detail in the contract, and should include stipulations that failure to meet a minimum amount of progress or specific project milestones will result in the repository taking possession of whatever digital files were created and making them openly accessible. This provides incentive to the publisher, assures that the investment is not altogether lost (the publisher will take a tax write-off), and ensures that collection digitization rights are not tied up indefinitely (if exclusive rights were granted).

The contract should also specify that scanner operators with demonstrated training and experience in the proper handling of original materials will be hired, and give the repository some role in selection of scanning personnel. Finally, the contract should also specify requirements for handling original materials. More details on digitization project planning and considerations can be found in materials from the Research Libraries Group, the Northeast Document Conservation Center, and NINCH.<sup>38</sup>

#### *E. Assigning Responsibility: Legal Terms*

The contract should assign to the publisher responsibility for compliance with all copyright and legal requirements applicable to the distribution of the digital collection and for obtaining all consents, permissions, licenses, and other instruments as may be necessary for such compliance. In return for accepting that responsibility, the publisher will need to reserve the right to eliminate

any items from the digital collection where, in the publisher's opinion, reproduction of the item would violate copyright law (or other laws of libel, obscenity, etc.), or in cases where clear copyright cannot be established. The publisher should also agree to indemnify the repository and hold it harmless from any costs, including attorney's fees, resulting from any claim of a violation of any rights of a third party arising out of the reproduction, publication, and sale of the digital collection.

#### *IV. Conclusion*

We have argued that digital dissemination of public-domain works of historic, literary, and artistic value increases the intellectual value of those materials and that repositories that license digitization rights to commercial agencies should seek the greatest possible availability for these works on the open Internet. As with copyright law, however, a balance must be struck between society's legitimate interest in maximizing access to and use of the work and society's equally legitimate interest in encouraging capital investment in digitization, dissemination, and long-term sustainability. By proactively working with vendors to contractually ensure access terms, distribution rights, and digitization standards, while at the same time recognizing the economic realities of the marketplace, repositories can provide digital access to their holdings of literary, historical, and artistic works in the public domain in a manner responsive to the broad needs of long-term open access.

## Endnotes

---

1. "Ownership of a copyright, or of any of the exclusive rights under a copyright, is distinct from ownership of any material object in which the work is embodied. Transfer of ownership of any material object, including the copy or phonorecord in which the work is first fixed, does not of itself convey any rights in the copyrighted work embodied in the object; nor, in the absence of an agreement, does transfer of ownership of a copyright or of any exclusive rights under a copyright convey property rights in any material object"—17 U.S.C. 202.
2. For example, *Making of America* is a joint project of the University of Michigan and Cornell University funded by the Andrew W. Mellon Foundation. Cf. <http://www.hti.umich.edu/m/moagrp/about.html>. The *American Memory Project* (<http://memory.loc.gov/>) is a collaborative project with the Library of Congress, partially funded by the Ameritech Corporation.
3. Examples include the microform and digital research collections offered by ProQuest/UMI, Primary Source Microfilm, and other companies.
4. For guidelines specific to administration of library rare book and special collections units, see Lisa Browar, Cathy Henderson, and Michael North, "Licensing the Use of Special Collections Materials," *RBM* 3, no. 2 (Fall 2002), 124-144.
5. Intellectual property laws that govern invention – patents – differ from intellectual property laws that govern expressions of ideas – copyright. This paper is concerned exclusively with copyright.
6. *Eric Eldred, et al., Petitioners v. John D. Ashcroft, Attorney General*, 239 F.3d 372 (2003), affirmed.
7. *Eric Eldred, et al., Petitioners v. John D. Ashcroft, Attorney General*, 239 F.3d 372 (2003), affirmed. (Breyer, J., dissenting).

- 
8. "Reclaim the Public Domain," <http://www.petitiononline.com/eldred/petition.html>.
  9. Lawrence Lessig, *The Future of Ideas: The Fate of the Commons in a Connected World* (New York: Vintage Books, 2002), 14.
  10. Lessig, 14.
  11. Lessig, 50.
  12. Quoted in Lessig, 204.
  13. Budapest Open Access Initiative, <http://www.soros.org/openaccess/read.shtml>.
  14. *Directory of Open Access Journals*, <http://www.doaj.org/>; accessed 30 October 2003.
  15. "Public Library of Science to Launch New Free-Access Biomedical Journals with \$9 Million Grant from the Gordon and Betty Moore Foundation," [http://www.plos.org/news/announce\\_moore.html](http://www.plos.org/news/announce_moore.html).
  16. Cf., among others, Stevan Harnad, "For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now," <http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm>.
  17. Budapest Open Access Initiative, op. cit.
  18. Public Access to Science Act, H.R. 2613, (108th).
  19. Lessig, 174, our emphasis.
  20. Lessig, 88.

- 
21. *The New Palgrave Dictionary of Economics and the Law*, ed. Peter Newman (New York: Stockton Press, 1998), s.v. “Network effects and externalities.”
  
  22. *A Framework of Guidance for Building Good Digital Collections*, November 6, 2001, <http://www.ims.gov/pubs/forumframework.htm>.
  
  23. “Stage 1, Diagnosis: Does Your Journal Meet Its Primary Goal -- To Serve Its Community?,” in: Scholarly Publishing and Academic Resources Coalition, *Declaring Independence*, <http://www.arl.org/sparc/di/stage1.html>.
  
  24. Steven Puglia, “Revisiting Costs” in: *The Price of Digitization: Cost Models for Cultural and Educational Institutions*, NINCH Symposium, New York, April 8, 2003, <http://www.ninch.org/forum/price.report.html>.
  
  25. *The Price of Digitization: Cost Models for Cultural and Educational Institutions*, op. cit.
  
  26. “Project Planning (Cost Models)” in: Humanities Advanced Technology and Information Institute, University of Glasgow, and the National Initiative for a Networked Cultural Heritage, *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, October 2002, <http://www.nyu.edu/its/humanities/ninchguide/index.html>.
  
  27. Steven Puglia, “Revisiting Costs” in: *The Price of Digitization: Cost Models for Cultural and Educational Institutions*, op. cit.
  
  28. Net receipts should be calculated based on payments received by the publisher from purchasers, subscribers, and distributors for access to all or any portion of the digital collection, adjusted for returns, customer and agency discounts and credits, and less sales tax, if any.
  
  29. *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, op. cit.

- 
30. *Building and Sustaining Digital Collections: Models for Libraries and Museums* (Washington: Council on Library and Information Resources, August 2001), <http://www.clir.org/pubs/reports/pub100/pub100.pdf>.
  
  31. D. Zorich, "Why the Public Domain is Not Just a Mickey Mouse Issue," NINCH Copyright Town Meeting, Chicago Historical Society, January 11, 2000, <http://www.ninch.org/copyright/2000/chicagozorich.html>; and Laura Gassaway, "When Works Pass into the Public Domain," September 18, 2001, <http://www.unc.edu/~uncldg/public-d.htm>.
  
  32. "Rights Management" in: *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, op. cit.
  
  33. Association of Research Libraries and others, *Principles for Licensing Electronic Resources*, <http://www.arl.org/scomm/licensing/principles.html>.
  
  34. International Coalition of Library Consortia, *Statement Of Current Perspective And Preferred Practices For The Selection And Purchase Of Electronic Information: Update No. 1: New Developments In E-Journal Licensing* (December 2001 update to March 1998 Statement), <http://www.library.yale.edu/consortia/2001currentpractices.htm>.
  
  35. *The Liblicense Model Licensing Agreement*, <http://www.library.yale.edu/~llicense/modlic.shtml>)
  
  36. *Creative Commons*, <http://www.creativecommons.org>.
  
  37. Digital Library Federation Benchmark Working Group, *Benchmark for Faithful Digital Reproductions of Monographs and Serials*, Version 1, December 2002, <http://www.diglib.org/standards/bmarkfin.htm>; Anne R. Kinney and Oya Y. Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, Calif.: Research Libraries Group, 2000); *Guides to Quality in Visual Resource Imaging*, (Research Libraries Group and the Digital Library Federation, July 2000), <http://www.rlg.org/visguides/>;

---

*Handbook For Digital Projects: A Management Tool for Preservation and Access*, Maxine K. Sitts, ed., (Northeast Document Conservation Center),  
<http://www.nedcc.org/digital/dighome.htm>; *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, op. cit.; Western States Digital Standards Group-Digital Imaging Working Group, *Western States Digital Imaging Best Practices*, Version 1.0, January 2003,  
[http://www.cdpheritage.org/resource/scanning/documents/WSDIBP\\_v1.pdf](http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf).

38. Anne R. Kinney and Oya Y. Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, op. cit.; *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, op. cit.