In Google's Broad Wake
Taking Responsibility for Shaping the Global Digital Library

BY RICHARD K. JOHNSON

It wasn't long ago that libraries confronted the challenge of how best to frame licenses for digital resources that were assembled by others and used by local patrons. Today the challenge is turned upside down as a growing number of research libraries enter into agreements that carry the collections the libraries themselves have built and nurtured out beyond their institutions and into the world.

Announcements of major projects to digitize library collections are becoming almost commonplace:

- Google has inked deals with a growing array of research libraries for mass digitization of public domain and in-copyright books in their collections.
- Fifty libraries, museums, technology companies, and others banded together in the Open Content Alliance (OCA) to undertake mass digitization of library collections.
- Microsoft has gotten into the game, both as part of OCA and via a plan to digitize public domain material provided by library partners for its Live Book Search portal.

And let's not forget the less-publicized arrangements with players such as those involved in the University of Michigan's Text Creation Partnerships: Early English Books Online with ProQuest, Evans Early American Imprints Collection with Newsbank/Readex, and Eighteenth Century Collections Online with Gale.[1]

Agreements involving institutions' born-digital resources also are part of today's mix. For example, ProQuest offers a service that provides online access to theses and dissertations from participating universities. Similarly, BioMed Central and several other businesses operate institutional repositories for universities.

The significance of such undertakings—mass digitization, in particular—has been widely recognized and discussed in the public media. But within academe, elation over the possibilities of popular access to many of the world's great libraries has been tempered by some unease about whether the deals go far enough in advancing or protecting institutional, scholarly, and public interests. Do these arrangements adequately prevent restrictions on use of public domain works? Do they advance the opportunities for resource discovery and computational analysis? Do they anticipate long-term preservation needs? Do they sufficiently value the substantial investment made by institutions to develop, organize, and maintain collections over time?

For public and private institutions alike—all tax-privileged or publicly funded institutions that steward cultural heritage and the scientific record—there are questions of whether contract terms fully serve the public interests and whether they are sufficiently transparent. The practical implications of failing to be open are illustrated

by the Smithsonian Institution's agreement giving Showtime Networks semi-exclusive rights to use Smithsonian resources for development of cable programming—a contract that was never released to the public. A controversy swiftly arose over the special advantage given to a single company, an arrangement that seemed to many to benefit the Smithsonian more than the public. (A Government Accountability Office investigation spurred by the controversy concluded the agreement did not hamper researchers' access.)

Given the growing base of experience in framing such agreements—and the prospect of more agreements to come—this is a good time to consider the issues of practice and principle that surround contracts specifying how library collections can be used by other parties and what can be done with the digital files created or managed by those parties.[2]

Sea Change

Digitization of library collections was hardly novel when it burst into the public consciousness with the first Google deals with libraries in late 2004. New, however, was the scale of investment Google promised to bring, the sweeping scope of their vision, the rapidity with which they sought to achieve this, and their bold approach to navigating potential legal obstacles.

The effect of Google Book Search, as the program is now dubbed, was to put library digitization efforts into an entirely new frame of reference. As a University of Michigan statement put it, "[we] had been digitizing works for years, but at a substantially slower pace than what was anticipated with the Google project."[3]

In more respects than just scale, Google's plan was fundamentally different from most of the digitization projects undertaken previously by libraries. Google brought more than money to the task; they also brought the imperatives of business—in particular, the need to protect their investment and erect barriers to competitive efforts. Google's business requirements meant restrictions on how the digital content would be used. As one blogger put it, "is it another case of 'Google can crawl everyone else's data, [but] no one can crawl Google's data'?"[4]

Library digitization has focused on preserving and curating high-quality digital transformations of out-of-copyright works, making full texts openly available for viewing, searching, crawling, downloading, and printing. The Google aim, on the other hand, was simply "to make offline information searchable online"[5] via Google's search engine. While Google's aim is not at odds with the needs or goals of the academy—indeed it promises to advance information sharing dramatically—Google Book Search isn't a perfect substitute for library digitization. And it does not necessarily anticipate the spectrum of opportunities and risks presented by digitization. For example, can anyone else build services that use the data—not just for indexing and access but also for other forms of computation? What will be the different treatments for works that are in copyright, orphaned, or in the public domain?

Donald Waters of the Mellon Foundation is among those who have voiced concern that

universities and libraries have been shortsighted about the potential uses of digitized materials:

> The discussion about commercially supported digitization projects has been primarily focused on keyword search to enhance discovery. The argument seems to be that the enhanced value of search and discovery is more than worth any negative conditions that might be imposed by commercial partners. However, there is much more to the usage equation that is of potentially great significance to the academy, and the danger is that research institutions may shortchange the academy and the public over the long term as they make their bargains with these firms. If we keep copyright issues to the side and focus only on public domain materials, digitization greatly enlarges the evidence base for scholarship. New kinds of analytical techniques are needed to exploit this new evidence for the advancement of knowledge, and research and teaching methodologies need to be developed to understand the evidence and train the next generation of scholars how to use it. Scholarly evidence, analytical techniques and disciplinary and pedagogical methods, which are so essential to the future of higher education in this country and elsewhere, can be greatly advanced—or seriously limited—depending on what rights the universities and libraries have to use, and let other institutions and scholars use, their copies of the digitized materials.[6]

From the perspective of Google and its competitors, keeping a rein on the digitized files enables them to protect their investment. Perhaps even more significantly, it ensures they don't lose control of in-copyright materials—an eventuality that brings with it the risk of illicit use and the corresponding certainty of Google being targeted by copyright owners' lawsuits. The prominence of this concern is implied by the relatively looser contractual controls over public domain works.

Despite any qualms, institutions were willing Google partners. Most described their motivations in similar terms:

- University of Michigan President Mary Sue Coleman described the partnership with Google as offering "three overarching qualities that help fulfill our mission: the preservation of books; worldwide access to information; and, most importantly, the public good of the diffusion of knowledge."[7]

- "'The digitization project furthers [the University of California's] mission,' said UC President Robert C. Dynes. 'It greatly expands our ability to give scholars and the public access to the kinds of information and ideas that drive scholarly innovation and public knowledge and discourse.'"[8]

- "The Bodleian Library's mission, from its founding in 1602, has been based on Sir Thomas Bodley's vision of a library serving the worldwide 'Republic of Letters', with the Library's collections open to all who have need to use them," explained Ronald Milne, Acting Director of Oxford University Library. "The Google Library Project in Oxford testifies to our ongoing commitment to enable and facilitate access to our content for the scholarly community and beyond."[9]

Showing a hint of scholarly skepticism, Harvard's then-president, Lawrence H. Summers, hedged a bit: "If this experiment is successful, we have the potential to provide the world's greatest system for dissemination."[10] In any event, Harvard and

others joined with Google knowing that their arrangements were not exclusive, thus protecting their fundamental interest.

Beyond Google

Google's 2004 announcement of five agreements was as close to a preemptive strike as could be imagined, but it was still only the beginning. Others have since joined the fray, libraries and businesses alike.

Apart from Google's controversial approach to in-copyright materials[11], the deals being struck have been largely framed by Google's first-mover advantage. As one library director reported privately, Microsoft initially proposed terms that were 180 degrees different from Google's, but as they've gotten down to brass tacks, the agreement has started to look a lot like Google's.

Ultimately, though, more experience together with competition for use of institutions' resources could mean better terms for the institutions. Libraries would do well to reject the notion that the contracts executed to date have established the norms for future agreements.

The Open Content Alliance, established in November 2005, pointedly embraced principles of open accessibility. Organizations that are contributing to the OCA have indicated their support of several broad principles:

1. The OCA will encourage the greatest possible degree of access to and reuse of collections in the archive, while respecting the rights of content owners and contributors.
2. Contributors will determine the terms and conditions under which their collections are distributed and how attribution should be made.
3. The OCA need not be obligated to accept all content that is offered to it and may give preference to that which can be made widely accessible.
4. The OCA will offer collection and item-level metadata of its hosted collections in a variety of formats.
5. The OCA welcomes efforts to create and offer tools (including finding aids, catalogs, and indexes) that will enhance the usability of the materials in the archive.
6. Copies of the OCA collections will reside in multiple archives internationally to ensure their long-term preservation and accessibility to all.[12]

According to OCA's Brewster Kahle:

> In essence, we want to get the rules right, to enable libraries to work with commercial sources, governments, etc., without having to hammer out separate agreements. Like the Open Source movement has done for software, we want the Open Content movement to do for institutions, to let them play a role. We want to define a way through the puzzle of who does what, to establish mechanisms for cooperation.[13]

The devil, of course, is in the details. The OCA Web site indicates that, "Generally, textual material will be free to read, and in most cases, available for saving or printing using formats such as PDF. Contributors to the OCA will determine the appropriate level of access to their content." However, it goes on to say, "OCA welcomes all efforts

to create and offer tools (including finding aids, catalogs, and indexes) that will enhance the usability of the materials in the archive."[14]

OCA sidestepped much of the legal risk taken on by Google by targeting only material that is either in the public domain or has the copyright holders' authorization. But, being funded by its contributors and donations, OCA lacks the deep pockets of Google, which is amassing its trove of scanned books at far greater speed.

Microsoft is one of the OCA participants and committed to fund the scanning of 150,000 public domain books.[15] They also are making deals on their own, including large-scale digitization of books from the British Library, University of California, Cornell University, New York Public Library, University of Toronto, and the American Museum of Veterinary Medicine. According to the October 2006 Cornell announcement:

> Microsoft will give the Library high-quality digital images of all the materials, allowing the Library to provide worldwide access through its own digital library and to share the content with non-commercial academic initiatives and non-profit organizations.[16]

But whether third parties will be able to develop services using these files is not stated and none of the Microsoft contracts have been made public.

The Microsoft digitization initiatives also are skirting thorny legal tie-ups by focusing on out-of-copyright scholarly material plus in-copyright books that are sent to Microsoft by publishers or authors for scanning by the company free of charge.

Yahoo is part of OCA, too, and will index the OCA content. It reportedly is "funding the digitization of an initial corpus of American literature collection that the University of California system is selecting."[17]

What's Ahead?

Given the broadening scope of book digitization projects now underway, it is reasonable to believe we will see more action in the months and years ahead to harness and exploit other kinds of knowledge assets. These might include:

- unique and unpublished materials in special collections of libraries and archives;
- course materials;
- theses, dissertations, and other student projects;
- images, video, sound, and animations;
- data sets;
- works contained in institutional repositories;
- government information; and
- technical reports and other gray literature.

Collaborations already reach beyond research universities. The New York Public Library was among the first wave of Google partners and also is working with Microsoft. The Library of Congress is now conducting a pilot project with Google. NASA is working with Google to give the public easy access to its mass of images and

data. The Biodiversity Heritage Library, a cooperative of botanical gardens and museums, is part of OCA. As the commercial sector digs deeper, it's a safe bet that attention will extend to other cultural heritage organizations and scientific research centers. And as digitization efforts turn increasingly to unique materials, institutional negotiating leverage may be greater.

Given the prospect of more deal-making, there is some popular sentiment among research libraries that the community would be well served by a set of principles to be upheld in—or at least provide a touchstone for—digitization or similar agreements. Such an assertion would increase the understanding and negotiating prowess of libraries and give weight to their positions.

Getting It Right

Whether or not the provisions of Google's agreements with libraries to date have cast the mold for other negotiations, they certainly are part of the market context. Fortunately, negotiations are already being guided by the strong values of the library culture and the importance of information sharing to the functioning of the academy. This influence is evident in the foundation of the Open Content Alliance—even if, in the OCA's effort to unite a diverse group of academic and commercial collaborators, they have made some compromises.

But the need for more than good intentions was acknowledged at an April 2006 Digitization Policy Workshop organized by the American Library Association's Office of Information Technology Policy (OITP) Advisory Committee and attended by representatives from the library and cultural heritage communities. A statement that emerged from the workshop observed:

> The Internet has brought the world of information to classrooms, homes and offices of people worldwide. For over a decade now, libraries and other cultural heritage institutions have been opening their rare and fragile collections to scholars and grade school children alike by digitizing these collections and making them available via the internet. In the last year, libraries have been moving from smaller digitization projects to mass digitization projects that will eventually make available whole collections, including millions of books. Funding agencies are supporting research and demonstration projects that aid libraries and cultural heritage institutions in better understanding digitization processes, web harvesting, tool development, and assessment and evaluation. All of this has taken place without a coherent body of policy to guide decision-making.[18]

In fact, over the past two years there have been a number of efforts—formal and otherwise—to articulate core library interests in digitization partnerships. Following are some examples that may help inform the agreements that have yet to be negotiated.

American Library Association

In the wake of ALA's Digitization Policy Workshop, the association's Office for Information Technology Policy constituted a Task Force on Digitization Policy, which drafted these "Principles for Digitized Content." The principles are to be presented to the ALA Council for discussion and potential adoption in 2007:

1. Digital libraries ARE *libraries*. The policies of the Association apply fully to digital libraries including the core values such as commitment to access, confidentiality/privacy, the public good, and professionalism.
2. Digital content, like other *library materials*, must be given the same consideration for collection development, ease of access, freedom of information, and preservation.
3. Digital activities and the resulting collections must be sustainable by libraries. *Sustainability* requires secure and ongoing funding, technology solutions that are appropriate to the longevity of the cultural record, and long-term management capabilities.
4. Digitization on a large scale requires *collaboration*. Collaboration enables the building of collections that support research, scholarship and information needs of diverse communities. Collaboration will require strong organizational support and promotion by cultural heritage professionals, their institutions, and their associations.
5. Digital activity requires ongoing *communication* for its success. The library and cultural heritage community must reach out to the public, to government, and to funding institutions with a clear and compelling message regarding the role of digital libraries and collections.
6. Digital collections increasingly address an *international* audience. These collections are part of a global information infrastructure that is not limited by geography.
7. Digital collections are developed and sustained by an educated workforce. Members of the cultural heritage professions must engage in *continuous learning* and be able to explore new technology, to work with new partners, and to reach new audiences.
8. Digital materials must be the object of appropriate *preservation*. Preservation activities require the development of standards and best practices as well as models for sustainable funding to guarantee long term commitment to these materials.
9. Digital collections and their materials must adhere to *standards* to maximize their usefulness. Standards must serve the broadest community of users, support sustainable access and use over time, and provide user functionality that promotes the core library values.[19]

The Andrew W. Mellon Foundation

In spring 2005, Donald Waters, Mellon Foundation Program Officer for Scholarly Communications, initiated an informal discussion with Daniel Greenstein of the California Digital Library and others on the need to frame a set of principles that would inform the negotiation of mass digitization agreements. A draft text, never promulgated, suggested that libraries bear "a special responsibility with regard to the future use and disposition of the collections they have assembled and managed in large part through public investment, protection, and other special affordances." The draft text argued:

> [M]assively digitized collections built in whole or in part on library holdings will uphold and support the public goods…where they can be contributed to massively digitized collections that are:
>
> • *persistent as whole collections* whose long-term management is governed by responsible third party archives empowered under specified business conditions to make collection content generally and widely available, for example, when they are no longer accessible via other means
> • *configurable* for example through the use of published and open standards that support and enable third-party service development and definition
> • *respectful* of the rights and prerogatives of copyright holders

- *attributable* in a standard way that makes it possible readily and transparently to identify the provenance of any item as well as the terms and conditions that govern its distribution and use
- *openly and freely accessible* at least where collection and item-level metadata are concerned
- *created through the use of non-exclusive agreements* that preserve for content providing libraries (perhaps after a time-limited period not exceeding six years) the non-exclusive worldwide and perpetual right to manage and distribute digital files that are created from their holdings.[20]

Francophone National Libraries

A high-level view emerged from a February 2006 meeting of six francophone national libraries, which endorsed these principles:

- Agreements with private sector partners to publish or digitize significant collections will be non-exclusive in nature, in both a legal and a de facto sense.
- Public access online to publicly owned resources will remain free. Information providers may develop and charge for value-added features but the source material should be accessible and free.
- The digital images will be prepared to a suitable preservation standard and maintained in the public sector with a commitment to long-term preservation and accessibility.
- As far as possible, online access will be multi-lingual and multi-cultural as appropriate for the source material.
- The integrity and authenticity of the original source material will be maintained and cannot be altered in the online environment.[21]

Committee on Institutional Cooperation

The perspective of a consortium of large libraries emerged from the Committee on Institutional Cooperation (CIC) library directors, who endorsed a set of principles in spring 2006. These principles relate to federating mass digitization efforts among the CIC libraries. Several pertinent principles are excerpted here:

- Digitization goals are for wide distribution, improved access methods (including the use of robust metadata), preservation of materials for the future, ability to use content in new ways that facilitate new forms of research, ability to leverage the investments of all CIC universities, and avoidance of unnecessary duplication of effort.
- Recognizing that standards are dynamic and evolving, the efforts for digitization, access, metadata, storage, and preservation will be based on generally accepted and widely adopted international standards.
- Resources combined through the federation will be "open access" throughout the CIC to the extent legally permissible, but it is understood that investments will be necessary for the development of the federated resources that might result in charges to users beyond the CIC.
- CIC libraries that sign agreements with vendors in order for those vendors to digitize material from library collections agree that such digitization will be done with the intent of encouraging open access and discouraging re-purposing of the content for sale by the vendor.[22]

OCLC's RLG Programs

In a draft text shared in November 2006 with the ARL Scholarly Communications Steering Committee, James Michalko of OCLC's RLG Programs division suggested several potential expectations of digitization partnerships—these, too, anchored in the notion of library collections as a public good, but recognizing that "expectations will be defined by situational needs." Following is an excerpt:

> Digitized works support the public good and scholarly use when they exhibit attributes that enable the creative use and re-use of collections built and managed as part of a common cultural heritage. Research institutions have a shared interest in securing the rights necessary to ensure that the products of mass digitization support or enable:
>
> • recombination
> • deeper access
> • more immediate access
> • more malleable text
> • derivative formats
> • links to related research, data, and primary sources
> • authenticity[23]

American Council of Learned Societies

In their 2006 report, *Our Cultural Commonwealth*, the American Council of Leaned Societies Commission on Cyberinfrastructure identified five "necessary characteristics of a robust cyberinfrastructure in the humanities and social sciences." Institutions' agreements to make their collections available on the Internet might usefully be evaluated against these five "necessary characteristics of a robust cyberinfrastructure":

> • It will be accessible as a public good.
> • It will be sustainable.
> • It will provide interoperability.
> • It will facilitate collaboration.
> • It will support experimentation.[24]

Cornell University

Principles and objectives are not the same, but a look at Cornell University Library's objectives in entering into a collaboration with Microsoft—extracted from an internal document entitled, "Why Are We Collaborating with Microsoft?"—illuminates both the motivations of the institution's leaders and the principles that suffuse their culture [headings added]:

• Advance global scholarship

> This partnership supports the Library's commitment to make its collections available to a wide audience in a high-quality digital format that meets users' sophisticated access and discovery requirements. We have a long-standing commitment to make our collections broadly available to support global scholarship.

- Support the work of the institution

   By partnering with Microsoft, we participate in content selection decisions to ensure that not only popular but also scholarly materials that support the CUL's academic programs are available in digital form on the Web.

- Enhance preservation

   By receiving all the digital scans back, we are not relying on a commercial entity to assume the preservation responsibility for either print or digital versions of this important scholarly content.

- Expand scale of online resources

   During the first year of the partnership, up to 170,000 monographs and serials will be digitized and be accessible on the web…. Cornell has been working for fifteen years to digitize significant collections, and this effort… enhances CUL's leadership role in transforming our collections.

- Improve return on investment

   Some of these materials are currently getting very low use. We believe that digitizing and making them available online will boost their use and will increase the long-term returns on our investment in the purchase, storage, and maintenance of Cornell's extensive print collections.

- Ensure flexibility

   The agreement with Microsoft is non-exclusive and the Library is free to join other partnerships with the same or different content in digitizing its collection. [25]

Advancing Global Interests

It is fair to assume that immediate local interests will mostly take care of themselves as institutions work out terms of partnerships that utilize their library collections. But, as the articulations above demonstrate, strategic, community-wide, and societal interests—ultimately ensuring a full flowering of the emerging global digital library—must be an explicit part of the negotiation and a legitimate object of applying available leverage. This obligation is of significance for public and private institutions alike.

The US National Archives and Records Administration (NARA) recently learned the hard way the danger of failing to adopt a set of principles to guide their negotiation of a digitization agreement with Google. Reports of the terms of an imminent agreement were greeted by much of the library community with "concern about the ownership—physical and legal—of both the underlying documents and the subsequent digital files."[26] As a consequence, NARA went back to the drawing board to develop a set of principles framing development of such NARA agreements.[27]

To aid institutions in conceptualizing the principles that may guide them, following is a simple clustering of some of the critical interests expressed above, colored by

perspectives gleaned from a sampling of interviews during fall 2006 with directors of six North American research libraries plus officials of three academic library organizations. This list is expressed not as principles but as broad issues and strategic interests against which potential agreement terms can be evaluated.

- *Exclusivity*: As a practical matter, exclusive agreements risk ceding crucial rights over use and reuse of digital files made from the collections an institution has assembled and supported at great expense. Future opportunities in a fast-changing environment are hard to anticipate, so any restriction of rights has both local and societal implications. From a local perspective, exclusive deals may inhibit the ability to obtain enhanced value from that investment in years to come. From a broader perspective, exclusive deals stifle creative uses of collections. Of course, exclusivity comes in many sizes and shapes, potentially defined by time, copyright status, geography, media/formats, markets, or other factors. However, institutions and society benefit most when exclusivity in any form is minimized.

- *Uses of the digital files:* The issues surrounding uses of the digitized works lie at the core of an institution's objectives in entering an agreement, which probably include expanding collection usage and supporting learning and new contributions to scholarship. As a practical matter, many of the uses envisaged in an agreement will need to be framed in light of the copyright status of works. While the rights to use public domain materials are wide open under statutory law—and should remain so in the environment framed by a contract—use rights with respect to in-copyright materials are less clear-cut and indeed may be interpreted by pending or future litigation. Within each of these two classes of work, the kinds of uses to be anticipated by a contract will likely touch not just upon what can be done with the partner's product offering, but also whether all or part of the digitized works can be rehosted by the source institution or others and whether services that interact with the database can be implemented by the institution or by third parties. While these services might very well include keyword searching and display of a result of specified scope (e.g., a "snippet" or something more), a forward-looking agreement will enable emerging computational applications such as text mining.[28]

- *Respect for the public domain and copyright*: Recognizing the rights of copyright holders is both a principled and a practical obligation, requiring that arrangements account for how information about ownership and use rights will be conveyed. There is also a danger here that, without explicit protections, restrictive contract provisions could supplant use protections granted under copyright law that are invaluable to scholarship. Academic institutions have an interest in preserving statutory fair use and educational exemptions and the status of public domain works. As a matter of principle and self-interest, agreements that introduce restrictions beyond those that apply to printed works should be resisted. In a partnership, there is also a danger that there will be disagreements about how the law will be interpreted. In their agreement with Google, the University of Wisconsin protected against such an eventuality by securing digital copies of public domain works scanned from their collection and the rights to use these. They plan to make the works available to the public under terms of the university's choosing.

- *Preservation:* All digitization projects have implications for preservation of works in

the collection. An objective of digitization may or may not be to create a digital preservation copy, which has a bearing on maintenance of the corresponding print collection. Indeed, in some cases the act of digitizing printed material endangers or destroys the printed copy. The issues go beyond standards to questions of selection, collection integrity, quality control, handling, loss or damage, accountability, return of artifacts, and potential multi-institution initiatives to prune collections of duplicate items. Regardless of whether digitization creates a preservation copy, any agreement should address the need for persistent access to the digitized works. A strategy is needed for redundant storage of files in geographically dispersed locations and there are also questions of who will migrate content and what will be needed over the long term to enable this—technically, financially, and organizationally.

- *Use of standards:* An institution's goals, especially with respect to use and preservation of works, are advanced when appropriate standards are employed by the entity that wishes to exploit the collection. The benefits of standards use may in some cases be greater for libraries than for commercial partners. Their use has implications for up-front costs but also maintenance costs and the need for refreshing of technology. Failure to apply existing standards can create intended or unintended exclusivity issues.

- *Quid pro quo*: Assuring adequate valuation of the library's contribution to a partnership is akin to a fiduciary responsibility to the institution that supports the collection. Measures of value might include royalties, cost savings, or acknowledgement of the library in products or services. But a key measure is the extent to which an agreement advances the community and societal benefits of faster, less burdensome, more open information sharing.

- *Transparency*: Because of their role in advancing and preserving knowledge, research libraries and their institutions (and most other cultural heritage institutions) enjoy enormous public trust. Indeed, in the US they are protected by statutory provisions such as the library privileges in copyright law and, in most cases, they benefit from being not-for-profit organizations under the tax code. It follows then that they are accountable to the public for the terms of the agreements and the extent to which the results of a deal will serve the public good. Their status brings with it a principled obligation to be as open and transparent as possible in their business dealings. Even as a purely practical matter, the public communication component of a deal needs to be taken into account. As Smithsonian Institution officials acknowledged in the wake of the public outcry over their confidential contract with Showtime, "[we] may not have done enough initially to inform others about how the contract would affect existing Smithsonian policies and procedures."[29]

Good Stewardship

The groundwork for the global digital library—what's been described as "a single liquid fabric of interconnected words and ideas"[30]—is taking shape in the broad wake of Google's ambitious book digitization program. In offering up their collections, research universities and other institutions have either overcome or guarded against what Ross

Atkinson described as their "morbid fear…of becoming mirror images of each other as we move toward a more unified collection."[31]

But like most changes of such scale and complexity, this virtual library will grow in iterative fashion, taking account of the interests of differing players, the attributes of differing kinds of works, the imperatives of differing legal circumstances, the vigor of emerging technologies and standards, and an evolving comprehension of what is possible and necessary.

Obviously, local circumstances and objectives will play a prominent role in most negotiations to make local resources available on the network. Inevitably, the possible will be balanced against the practical. But the interests of each individual institution are best served when all are negotiating vigorously for agreements that maximize the full potential of networked digital libraries. If the ultimate mission of the academy is to advance knowledge, promote teaching and research, and address the challenges that face our society, then resources stewarded by libraries should be shared as widely as possible.

As additional resources are targeted for dissemination on the Internet, libraries and their institutions should negotiate strategically. They should make full use of whatever leverage they have to ensure the global digital library is open and dynamic.

# # #

[Sidebar:]

Negotiation Checklist

Here are some of the issues to weigh as your institution considers entering an agreement involving digitization or other business exploitation of its collections. For many questions below, the answer will differ for public domain vs. in-copyright material. These points only begin to outline the multiple layers of questions that institutions should ask before they enter significant agreements. Moreover, there are many additional practical issues to plan for—such as warranties and indemnifications, term and termination, cost apportionment, scheduling, and workflows.[32]

Exclusivity
• Are you able to enter similar agreements with other parties?
• Are there any restrictions on this right?
• Where exclusivity may be justified, what is the maximum period over which it applies?
• Are exceptions needed? Are there other ways exclusivity can be limited?

Uses of the digital files
• Does the agreement distinguish between uses of digitized public domain content vs. in-copyright content?
• Will your institution be provided a copy of any transformation of works from your collection (along with associated metadata and image coordinates)? In what form will this be provided?
• What rights does the institution have to use the digital files and with whom may they be shared—members of your institutional community? Your consortium? Any user of the Web?
• What provisions are there for the library to manage the files (e.g., take down content when warranted, reformat, add enhancements)?
• Under what terms will end users have access to content? For how long are these terms guaranteed?
• Under what circumstances will users have access to different views of works (e.g., "snippets," "limited preview," full page views, downloadable PDFs, etc.)?
• Will works be attributable and employ durable identifiers?
• Will works be open to computational analysis (e.g., text mining)? Under what circumstances?
• Is there a right for third parties to create services (such as discovery or presentation tools) without unreasonable limitation?
• What uses are prohibited?
• Will restrictions lapse after a given period of time?
• Are there provisions for any change in the library's rights if the other party becomes insolvent or no longer makes the files available?

Respect for the public domain and copyright
• Are copyright restrictions being inappropriately extended to public domain works?
• Who decides whether a work is in the public domain or not?
• Will treatment of currently in-copyright works change when they enter the public domain in the future?

- Are fair use and educational protections of the copyright law preserved?

Preservation
- Are the formats being used appropriate for content migration over time?
- Will characteristics of the original publication (such as color or complex graphics) be replicated?
- Will original items be returned to the collection?
- Will destructive processes be used in digitization? How will fragile materials be handled?
- Who will select items to be included?
- What arrangements are there for quality control?
- What steps will be taken to ensure the integrity and utility of digital files? How will this be funded? Who is responsible?
- Will copies of digital files be kept in multiple geographically dispersed locations?
- Will the works persist as an identifiable collection?

Standards
- What standards will be adhered to with respect to image quality, metadata, and file formats?
- Will standards employed allow third parties to develop services?
- Will specified standards be non-proprietary and sufficiently flexible to accommodate the changing environment?

Quid pro quo
- Is the library/institution acknowledged in all uses granted under the agreement? What forms will this take?
- Will the library/institution receive royalties or other payments? (Although it is not a contractual issue, consider whether this will be credited to the library's budget.)
- Are research, teaching, and learning advanced sufficiently by the agreement?

Transparency
- Have you retained the right to make a copy of the agreement available to the public?
- How will you communicate to your institutional community and to the public about your goals in entering the agreement, how the goals will be effected, and the steps you have taken to advance the interests of your community and the public?

# # #

---

[1] These arrangements have their roots in the earlier microfilming of library collections by a multiplicity of companies under diverse contractual arrangements. The history of such deals with micropublishers may offer lessons for the present, but it is important to recognize that some of the fundamental issues at stake in the networked digital environment are quite different.

[2] Beyond the scope of this article, but of no less significance, are the strategic implications for libraries of mass digitization. An expanded understanding is needed of looming questions, such as: How does mass digitization affect scholarship? To what kinds of uses do we envision digitized resources will be put? What roles will libraries play in this new environment? What services can libraries develop to support emerging and evolving uses? Can libraries leverage

digitization to manage print holdings differently? How can resource management responsibilities and domain expertise be brought together?

[3] University of Michigan News Service, "Google Library Partnership," November 14, 2005, http://www.umich.edu/news/index.html?BG/google/statement

[4] Comment on "Google 'Showtimes' the UC Library System," Television Archiving Blog, comment posted August 13, 2006, http://www.archival.tv/2006/08/13/google-showtimes-the-uc-library-system/.

[5] Google Press Center, "Google Checks Out Library Books," December 14, 2004, http://www.google.com/press/pressrel/print_library.html.

[6] Donald Waters, e-mail message to author, January 10, 2007.

[7] Kevin Bergquist, "Google Project Promotes Public Good," *University of Michigan University Record*, February 13, 2006, http://www.umich.edu/~urecord/0506/Feb13_06/02.shtml

[8] University of California Office of the President, "UC Libraries Partner with Google to Digitize Books," August 9, 2006, http://www.universityofcalifornia.edu/news/2006/aug09.html.

[9] Google Books Search Library Partners, http://books.google.com/googlebooks/partners.html.

[10] "Harvard Libraries and Google Announce Pilot Digitization Project with Potential Benefits to Scholars Worldwide," *Harvard University Gazette*, December 14, 2004, http://www.hno.harvard.edu/gazette/daily/2004/12/13-google.html.

[11] For more on this rich topic, see: Jonathan Band, "The Google Print Library Project: A Copyright Analysis", *ARL: A Bimonthly Report*, no. 242 (Oct. 2005): 6–9, http://www.arl.org/bm~doc/arlbr242google.pdf.

[12] Open Content Alliance, "A Call to Participate in the Open Content Alliance," http://www.opencontentalliance.org/participate.html.

[13] Barbara Quint, "Open Content Alliance Rises to the Challenge of Google Print," *Information Today*, October 3, 2005, http://www.infotoday.com/newsbreaks/nb051003-2.shtml.

[14] Open Content Alliance, "FAQ," http://www.opencontentalliance.org/faq.html.

[15] Gary Price, comment on "Microsoft Announces MSN Book Search; Joins Open Content Alliance," SearchEngineWatch Blog, comment posted October 26, 2005, http://blog.searchenginewatch.com/blog/051026-000127.

[16] Cornell University Press Relations Office, "Cornell University Library Announces Partnership with Microsoft That Will Allow Global Access to Its World-Class Resources," October 17, 2006, http://www.news.cornell.edu/pressoffice1/Oct06/library_microsoft.shtml.

[17] Brewster Kahle, comment on "Announcing the Open Content Alliance," Yahoo! Search Blog, comment posted on October 2, 2005, http://www.ysearchblog.com/archives/2005_10.html.

[18] American Library Association, Task Force on Digitization Policy, "Background," http://www.ala.org/ala/washoff/contactwo/oitp/digtask.htm.

[19] American Library Association, Task Force on Digitization Policy, "Principles for Digitized Content (Draft)," http://www.ala.org/ala/washoff/contactwo/oitp/digtask.htm#prin.

[20] "Upholding the Public Goods of Massively Digitized Library Holdings," unpublished manuscript obtained by the author from Daniel Greenstein on October 3, 2006.

[21] Ingrid Parent, August 10, 2006, e-mail message obtained by the author.

[22] "Federating Mass Digital Efforts in the CIC Libraries, Endorsed by the CIC University Librarians, Spring 2006 (V.1.2)," unpublished manuscript obtained by the author.

[23] James Michalko, "Mass Digitization Partnerships—Rights and Responsibilities of Research Institutions," unpublished manuscript obtained by the author on October 22, 2006.

[24] Marlo Welshons, ed., *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences* (New York: American Council of Learned Societies, 2006), http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf.

[25] "CUL Large-Scale Digitization Initiative," unpublished manuscript dated September 2006, obtained by the author and quoted with permission.

[26] Association of Research Libraries, American Association of Law Libraries, American Library Association, Special Libraries Association; letter to NARA; July 6, 2006; http://new.arl.org/bm~doc/naragoogleletter.pdf.

[27] NARA subsequently announced a non-exclusive agreement with Footnote, Inc. to digitize selected records from the Archives' holdings. See: http://www.archives.gov/press/press-releases/2007/nr07-41.html.

[28] A useful exploration of computation access to scholarly literature is found in: Clifford A. Lynch, "Open Computation: Beyond Human-Reader-Centric Views of Scholarly Literatures," in *Open Access: Key Strategic, Technical, and Economic Aspects*, ed. Neil Jacobs (Oxford: Chandos Publishing, 2006), 185–193, http://www.cni.org/staff/cliffpubs/OpenComputation.htm.

[29] Jacqueline Trescott, "Smithsonian Deal with Showtime Passes Muster," *Washington Post*, December 16, 2006, C1, http://www.washingtonpost.com/wp-dyn/content/article/2006/12/15/AR2006121501894_pf.html.

[30] Kevin Kelley, "Scan This Book!" *New York Times*, May 14, 2006, http://www.nytimes.com/2006/05/14/magazine/14publishing.html?ex=1305259200&en=c07443d368771bb8&ei=5090.

[31] Ross Atkinson, "Six Key Challenges for the Future of Collection Development," *Library Resources & Technical Services* 50, no. 4 (October 2006): 245. It is worth noting that Atkinson concluded individual libraries will never lose their competitive differentiation because "[e]ach has vastly different resources—not only financial, but also human and creative resources, including different visions and values." However, he cautioned that libraries' collection services would likely face new competitive forces.

[32] For a useful look at relevant models and elements of business relationships between educational institutions and the commercial sector, see Peter B. Kaufman, "Marketing Culture in the Digital Age: A Report on New Business Collaborations between Libraries, Museums, Archives, and Commercial Companies", August 25, 2005, http://www.intelligenttelevision.com/MarketingCultureinDigitalAge.pdf.