Good Terms - Improving Commercial-Noncommercial Partnerships for Mass Digitization

A Report Prepared by Intelligent Television
for RLG Programs, OCLC Programs and Research

Peter B. Kaufman and Jeff Ubois
Intelligent Television

Executive Summary

In 2007, OCLC Programs and Research engaged Intelligent Television to study the partnership agreements between cultural institutions and for-profit companies for the mass digitization of books and other media. This report presents the findings of that study.

Libraries have been digitizing portions of their collections for more than twenty years, but recent opportunities to work with private partners, such as Google, Microsoft, and others, on mass digitization has opened up possibilities that were unimaginable just a few years ago. Private funding, commercially developed technology, and market-oriented sensibilities together may generate larger aggregations of digitized books far sooner than the library community had dreamed possible. There are many efforts underway to assess various aspects of these partnerships; this paper focuses on the terms in mass digitization agreements that affect research-community-centered outcomes.

The libraries and other cultural institutions that private companies first approached saw significant potential in these overtures; they were diligent in seeing that near-term local needs were met. Only when it became clear that a significant number of these partnerships were underway did the library community as a whole begin to think about the overall impact of these business relationships on the future of scholarship.

When we fantasize about that future, we imagine a single way to search all digitized books, journals, and other media; a combined index of all the full texts that will enable research that is otherwise impossible; a variety of tools to facilitate working with these materials; and the ability to create personal subsets of materials for deeper investigation. These goals cannot be realized if each commercial partner puts a fence around the materials that it digitizes and requires its institutional partners to fence in their copies as well.

This report is no substitute for sound legal advice. Attorneys are key players in these negotiations. The institutions they represent will want to inform their counselors of what they hope to get out of their partnerships (and what they hope to avoid), so that the attorneys can negotiate toward those ends. If, before they begin those discussions, institutions define certain desired outcomes, think through the effect of any likely compromises, and agree on the walk-away points, there is a much greater possibility that the results of these endeavors collectively will have a positive impact on scholarship.

While the private partner is investing significantly in these efforts, the contributions of the library partner should not be undervalued. Libraries are contributing not only individual books and other materials to these partnerships, but centuries of dedication to the selection, description, conservation, and curation of their collections. That said, it is also imperative that libraries understand the motivations and requirements of their business partners.

Based on our consultations with staff involved in more than twenty of these partnerships and many other interested parties, we believe that libraries should be prepared to negotiate intensively in pursuit of the following objectives:

- **Limited confidentiality** – Libraries should be sensitive to private partner needs to protect business and technology secrets, but insist on their own right to discuss aspects pertaining to their broader community. These deals involve some of the most complex decisions libraries will face, they can be improved through consultations with others, and libraries do not want to be in the position of having to refuse advice to peers who seek their guidance.
- **More complete deliverables** – Librarians must have input into the specifications of quality and formats and be...
on these events:

research community.

courted, a sense of urgency set in.

and months that followed, as it became clear that Google was serious and that other institutions were being

staff reacted to their offer with disbelief.

When Google executives separately approached the initial five libraries about digitizing their holdings, at first many

in law.

motivations of the institutions can be matched by good terms in the documents that enshrine them in practice and

private partnerships and to see if there is room for improvement in the key operating principles—to see if the good
digitization fall outside the scope of this report.

For the most part, details about the intricacies of copyright law, preservation, and the technical specifications of

digitalization fall outside the scope of this report. Rather, our effort is an attempt to understand the extraordinary work

that has already been done by cultural and educational institutions and private enterprise in crafting these public-
digitization archives. Each clause in each contract represents in its small way this intersection—this give-and-take—of
business strategy and institutional mission.

In light of these options and the momentum behind more than twenty-five Google partnerships and several other

partnerships with a variety of commercial entities, OCLC Programs and Research engaged Intelligent Television to
assemble the publicly available agreements from commercial-noncommercial mass digitization partnerships; collect
and organize commentary on these public-private deals; analyze in some detail the publicly available agreements
and assess the deals where the contracts are not publicly available; coordinate a discussion involving
representatives of institutions active in these partnerships (held in New York City on June 11, 2007); and prepare
this document—presenting the findings in a form that might be described as a mix of specific recommendations,
general best practices for deal-making, and broad statements of principle.

For the most part, details about the intricacies of copyright law, preservation, and the technical specifications of
digitization fall outside the scope of this report. Rather, our effort is an attempt to understand the extraordinary work

that has already been done by cultural and educational institutions and private enterprise in crafting these public-
private partnerships and to see if there is room for improvement in the key operating principles—to see if the good

motivations of the institutions can be matched by good terms in the documents that enshrine them in practice and

in law.

When Google executives separately approached the initial five libraries about digitizing their holdings, at first many

staff reacted to their offer with disbelief. One librarian reportedly wondered what they were smoking. In the weeks

and months that followed, as it became clear that Google was serious and that other institutions were being
courted, a sense of urgency set in. In retrospect, many decisions may have been made in that hurried and
competitive context without sufficient attention to the mission of each institution and to the overall welfare of the
research community. One librarian who had been party to such decisions later reflected lyrically, even Homerically,
on these events:

We were approached singly, charmed in confidence, the stranger was beguiling, and we embraced. For the love of selfish confidence, we spoke neither our fortune nor our misgivings with our neighbors or our friends. We felt special, invited to loud weddings on far away islands of adventure; in the quiet we may wonder if we were given broken jewelry.

We could have chosen to question the worth of the marriage under the terms offered. We might have chosen to hold off the sequined cloaks of confidence that wrapped our suitor's gifts. The glamour of it! Yet we knew there were other wives already, and we might have joined them in union, consulted openly, wondered what would be best for all instead of merely ourselves; but our concerns were narrow. In our selfishness, and wrapped in the fears we were given, we re-wrote and redefined our aims, misplaced our responsibilities, allowed the light and glory of the ideal to suffuse its glow over the
Now that the fever is abating, libraries can take a step back, imagine the world that they want to live in, and set forth a vision of that better world—defining it in terms that provide a place for the commercial sector without ceding to that sector the determining role.

This report sets forth objectives of the businesses with which libraries are partnering, describes a number of the aspirations and basic hopes of the cultural and educational communities, lists some practices that institutions might consider as they enter into future partnership agreements, and tries to conjure up a sense of how cultural institutions might act as a community rather than in isolation.

In collecting, distilling, and publishing this information, we have endeavored to present our effort as part of an evolving set of conclusions. This report reflects our thinking as of October 2007, and will join other contributions from the community as these deals (and commentaries on them) continue to proliferate. Already, related efforts are producing results, ranging from the Council on Library and Information Resources investigation of the preservation aspects of these partnerships, to the Digital Library Federation’s exploration of the issues as they relate to moving image collections, to work highlighting the value of open content.

The recommendations here have varying levels of utility, depending on where the reader is in the negotiation process. Some, we hope, may be useful for those just considering how to enter into mass digitization. Some, we hope, may be useful for those already in the thick of negotiations. And some, we hope, may encourage those with experience in this process to share what they have learned with the community. We hope that representatives of cultural institutions that are involved in public-private mass digitization will review the current trajectories of those efforts, establish new guidelines and principles, and explore the role they might play as the community strives toward optimal outcomes.

**Business Partner Perspectives**

The authors of this paper have been writing about and attempting to improve business partnerships from the perspective of nonprofit noncommercial cultural and educational institutions—Columbia University, UC Berkeley, the Library of Congress, the Internet Archive—for several years. That said, we have also been active in the commercial world, working for public and private companies in media and technology as well as startups funded by venture capital and angel investors. Before assessing the rewards and risks associated with partnerships for mass digitization, we believe it may be useful to consider the perspectives of commercial enterprises as they appraise the potential for working with cultural and educational institutions. Appreciating the priorities of business partners helps to compare and contrast their motivations, values, interests, and preferred outcomes with those of the noncommercial institutions seeking to build the digital future through these joint ventures. We therefore recommend that those in the noncommercial sector:

1. Recognize that the business models of companies in digitization deals today are often fresh iterations of longstanding business models espoused by companies who have long been active in the field, such as ProQuest and Readex NewsBank. Those arrangements often resulted in copies of original library materials being provided to their partner libraries (and revenue generated from the sale of those products was sometimes shared with the library that contributed the original materials). 4
2. Recognize, equally, that some of the business models are entirely new, reflecting the fact that these commercial enterprises are now producers or co-producers of screen-based media in the digital age and much of that media is now advertiser-supported, provided by paid subscription, or on-demand.
3. Recognize that commercial companies seek returns on their investments in mass digitization based on economic calculations including:
   - a. effects on near-term revenue;
   - b. effects on closing future deals that in turn may bring in additional future revenue;
   - c. effects on corporate profit;
   - d. effects on closing future deals that may bring additional profit; and
   - e. effects on company valuation
   - and that these calculations may tend to be more short-term than those of cultural and educational institutions.
4. Consider that the investment banks and management and strategy firms that appraise and influence the valuation of business partners are paying increasing attention to new ways of measuring value in the online world. Rendering content searchable, or what some have called “computably competitive,” may be as good an investment of resources, if not better, than simply making more content available. 8
5. Recognize that private companies often need a competitive edge in return for their investment. They desire to accumulate content that their competitors do not have; they want some measure of exclusivity to allow their investments to return value; and they want to protect their plans and approaches to technology and media.
6. Recognize that businesses see cultural and educational institutions playing a critical role in the $12 billion annual business of online search—not only as a customer and as a gateway to these online services, but also as a potential competitor.
7. Finally, recognize that while commercial companies are bringing to negotiations and business deals their billion-dollar valuations and hundreds of millions of dollars in capital to support digitization, they may not recognize that cultural institutions themselves represent cumulatively billions of dollars of investment, based on the value of their assets and decades (if not centuries) of collecting, curating, and preserving physical copies of these works.

These last two points underpin our approach to improving the equity of these partnership deals by encouraging both sides to better evaluate what they and their prospective partner bring to the table right at the start.

**Partnership Rewards and Risks**

The digitization agreements concluded over the last three years represent some of the most exciting new possibilities for libraries, museums, and archives in decades. Their potential to make collections far more accessible; to reach their patrons, patrons at sister institutions, and other users around the world; and to support traditional and innovative applications from research to teaching to publication is core to the present and future mission of the library community.

Partnership agreements can be evaluated in many ways and by many criteria, but perhaps the most utilitarian approach appraises the agreements by the light of what they could achieve—or what, if they had been structured and written differently, they could have achieved—for the educational community on the one hand and the business community on the other. In assessing how the negotiating partners balance the rewards that each side will earn (more digitized content; more ad revenue) and the risks each side takes (a de facto exclusive business arrangement; a significant cash investment in a library), one might explore how that balance can continue to be improved as deal-making in this field continues.

With this approach in mind, we believe it might be possible to inspect these agreements much like a mechanic might inspect a car every year—not checking every screw and plug and piston, but focusing on six aspects of agreement negotiation and execution. These areas are:

1. each side's policy regarding sharing information related to the partnership;
2. the deliverables and their ownership;
3. use and reuse rights to those deliverables;
4. rights to combine content with that of other institutions and organizations;
5. access and privacy issues related to usage data; and
6. details surrounding the conclusion of an arrangement.

**1. Transparency, Confidentiality, and Non-Disclosure Agreements**

Many members of the library community who have visited the GooglePlex in Mountain View are struck that even at the registration desk the initial, basic act of visitation is governed by a non-disclosure agreement. Being too open with business processes can be fatal for a company in the information business; corporate responsibility requires protecting confidential information. By contrast, cultural institutions draw their financial and moral support from a public that expects transparency in their activities, ranging from their materials acquisitions to their business deals. These institutions have established traditions of consultation and information exchange with colleagues in the field, as well as with their user communities.

Most mass digitization partnership negotiations have started with a request from the prospective commercial partner for a confidentiality agreement—often commercial partners have asked cultural institutions to treat even the act of considering a partnership as a secret. Today, several significant contracts, including Google agreements with some university libraries and the Smithsonian Institution's agreement with Showtime, remain secret, governed by a nondisclosure agreement (NDA) even as other, often quite similar, agreements are made available to the public, either willingly by the partners or by laws requiring the public disclosure of activities of state institutions.

While it may not be an option for libraries to refrain entirely from signing NDAs, those institutions that sign NDAs prohibiting cooperation with their peers will likely find themselves at a disadvantage. The sometimes virulent backlash against non-profit cultural institutions entering into secret business deals suggests that the rewards may be greater to both parties if deals are open to public or community involvement before the negotiations are concluded.

The decision regarding signing an NDA is not binary; NDAs come in different forms and have different effects. NDAs that limit discussion between librarians, curators, and archivists over the merits of proposed partnerships are fundamentally different from NDAs that limit disclosure of technical processes unique to a commercial partner. There is a difference worth appreciating between an NDA involving an exchange of cash for products or services and an NDA that involves commercial use of library holdings maintained on behalf of the public.

Though libraries and other public institutions frequently sign NDAs with provisions that protect a company's information about their practices and finances; provisions sealing information about partnerships in which public goods are the exchange medium are different. The terms in an NDA should be negotiated.
Many of those we interviewed observed that cooperation among libraries could provide negotiating power, but lamented that competition between libraries makes anything like collective bargaining very difficult. Though institutional competition may have induced and may still induce some libraries to sign away their rights to discuss their activities with their peers, these gradations in confidentiality are worth noting.

Recommendation: Open the Discussion, Consult with Peers

Cultural institutions should consider their own responsibilities for ensuring transparency in their operations and should negotiate more aggressively with their commercial partners over what provisions in non-disclosure agreements are truly necessary. Commercial enterprises should understand that in the library world, NDAs can jeopardize the success of the joint venture by arousing suspicion about otherwise good agreements. We recommend examining whether the confidentiality believed essential for various aspects and in various stages of the deal-making process is really necessary—and whether it is acceptable, given the cost of losing access to the advice of peers and losing the ability to speak openly about their experiences.

The National Archives recently opened their general plan for digitization\(^6\) as well as for a prospective business deal\(^7\) to public comment, thereby involving their communities in the process, potentially improving the results, avoiding a public relations catastrophe, and earning admiration for their openness—a useful precedent for others.

Limiting the secrecy imposed by nondisclosure agreements and other confidentiality provisions to business practices and technical processes will allow future negotiations the benefit of commentary by others in the library community and the public. Ideally, as these deals proliferate and both sides see the value of the growing digital corpus, confidentiality provisions will not extend to information about the forms (text and/or images) or the quality of the deliverables, nor to general information about the content (the quantity, rights status, subject areas, or media types) that is being contributed by the institution.

2. Ownership: Creating, Receiving, and Retaining Data

If, in certain ways, the Google and other agreements we discuss here are successors of older microfilming projects by ProQuest and others, then some comparison can be made between the business deals in these microfilming projects from the 1970s and 1980s and the business practices of their digital descendants. Microfilming companies were willing to agree in their negotiations not only to handle the original materials (often brittle material like 18th-century newspapers) with utmost care, but to provide the contributing institution with complete master and access copies of the film or fiche, as well as assessments regarding rights and the licensing opportunities for this material.

Complete copies of the digital files created in today's digital projects could also be provided to institutions, but a review of the agreements indicates that this is not often the case—frequently what is returned to the institution is just a portion of the outputs.

The National Archives 2007 digitization principles address this topic: “Partners shall provide NARA without charge a full set of the digital copies produced by the partnership. These copies shall adhere to NARA's technical specifications. Ultimately, NARA will have unrestricted ownership of these copies, including the right to make these copies freely available online.”\(^8\)

Just as the commercial partner will ask the institutional partner to respect its rights to the property (intellectual and physical) in which it has invested, the commercial partner also needs to be reminded of the long value chain behind public domain works, as a way of explaining how and why these need to remain in the public domain and not be subject to what some legal scholars call re-enclosure. Often the private partner will agree to public access (for free or for fee) to the public domain materials, but provide only very limited access to other materials. Re-enclosure happens when access to works otherwise in the public domain is limited by locking it up in new ways. Under U.S. copyright law, digitization in itself does not create a new work that could be copyrighted, but some of the digitization contracts now in place have a similar effect. In one extreme example, 19th-century government documents were being re-enclosed by one of the business partners.

The National Archives' digitization principles are instructive on this point, too: “Public access to publicly owned resources will remain free. Partners may develop and charge for value-added features, but access to the digital copies ultimately should be readily accessible and free.”\(^8\)

While the commercial partner, in order to recoup its investment, understandably does not want to allow its competitors unfettered access to the digitized content, it is also limiting access by the very users on whose behalf the library has amassed those materials—and who the library hoped to benefit by entering into the partnership. These differences can be addressed by allowing the commercial partner limited exclusivity for a period of time, after which the library will assume unlimited ownership.

Finally, it's worth noting that there is the potential “Swiss cheese” effect that may occur when only selected portions of a library's collections are scanned. Several interviewees noted that as scanning projects begin to emphasize non-duplication (i.e., to avoid scanning books previously scanned elsewhere), the sets of scans passed...
back to each library will be incomplete representations of library holdings, while the collections held by the commercial partners will include a union of all the diverse sets of scans.

Recommendation: Collect and Retain All Digitized Data

We recommend that agreements spell out in greater detail than they have done to date which deliverables will be provided, at what point in time, and in what condition. Libraries should be able to opt to receive any of the following: master images, access images, OCR texts, the coordinates that map the text to the images, records with links, metadata describing all of the files that are associated to a record and their sequence, technical metadata, and identification of items that were rejected (for the library's later attention). Further, the institution should have a role in specifying quality requirements and in reviewing the output to ensure that those requirements have been met. Libraries should also receive usage data (see section 5 below).

The contributing institution should ensure that they ultimately will own the digital copies without limits on access other than the copyright inherent in the originals.

3. Use and Reuse: New Services and Applications

The value of digitization partnerships can transcend expanded access to individual books by offering new functionality, combinations, and collaborative possibilities. Several uses of a large aggregation of texts seem particularly promising:

- **New ways of combining content.** Books in digital form can be mixed and recombined in diverse ways. Allowing instructors to assemble and redistribute course readers (i.e., compilations of selected materials) is a very promising use for digital files, but it is possible to go further. For example, all the images in a certain genre of books could be extracted and compiled to form a new work that highlights a particular art or science, or all the geographic locations mentioned by an author might be plotted on a map.

- **Development of new approaches to search.** Google is unlikely to be the last word in information retrieval. Digitized collections should be open to automated access by researchers developing new search tools.

- **Print-on-demand.** New printing technologies have decreased the cost of printing a book to less than a penny a page. By some estimates, this is less expensive than re-shelving a book in a large library. Print-on-demand services will be in greater demand as the number of books available to print increases.

- **E-book readers, the $100 laptop, and other access devices.** Among the most innovative possibilities are those intended to assist readers with conditions such as low vision and dyslexia; to allow children to explore books more intuitively; and to allow connections over relatively low bandwidth networks to reach users in developing countries. Broad access and the ability to download public domain books will increase the reach of these devices.

- **Annotation and other services based on usage and citations.** Margin notes have been a part of the book culture for centuries. Digital books can take advantage of annotation systems capable of handling notes from many people. More passive approaches involving analysis of usage patterns may help identify the most relevant texts, within a book or across books.

- **Translation and analysis.** Statistical machine translation work and linguistic analysis can be advanced by comparing large collections of digitized text. The availability of large collections of classic world literature seems likely to serve as fertile ground for research in this area. These technologies improve with the size of the text corpora being compared.

- **Access to special collections.** Though special collections are not included in most high-volume projects, the benefits of digitizing rare books and a variety of non-book materials will lead to a growing number of projects in this area. Special collections are often difficult to share outside a given institution, but such collections can be of great relevance to widely dispersed researchers.

In addition to these uses, libraries, museums, and archives stand to benefit from combining the output of digitization efforts. Why digitize a book, for example, when it has already been digitized elsewhere? By pooling the content, institutions can reduce costs and avoid duplication of effort. Access to pools of digitized books could effectively give every library user access to an enormous collection.

Many libraries do not have the funds to build the technical infrastructure needed to support large digital collections. As a result, the libraries rely on their business partners, not just for digitization, but for access services as well. These access services need to address a range of technical challenges (e.g., stable URLs, analysis of usage patterns, access controls on copyrighted works). Unfortunately, a complex array of institutional, technical, and legal barriers resulting from current agreements will make it difficult for libraries to contribute to the process of envisioning new applications, enhancing institutional cooperation, and creating an improved mix of services.

The balkanization of internal curatorial processes within many institutions—owing to a mix, perhaps, of corporate culture, resource constraints, separate specialties, and various personalities—often results in certain parts of an institution not knowing what another part of the institution is doing. A greater sense of the potential uses and reuses of material needs to be developed—along with an appreciation of what contract language will do, if librarians are not careful, to limit such uses and reuses.
Technical barriers to achieving the full potential of text digitization are becoming more apparent. Scholars will increasingly wish to link to particular passages within digital books. This ability presupposes that URLs will remain stable and requires a deep link to a specific location within a book. Stable URLs in other types of media, such as sound recordings and moving images, are equally important and should be a consideration in any negotiation. More open content will encourage nonproprietary approaches to these and similar challenges.

Despite the noble objectives stated by libraries and their commercial partners, language in contracts that restrict what users may do with content may work against the interest of the community over the long term.

**Recommendation: Preserve Traditional Freedoms to Support Research and Scholarship**

Library negotiators should work to ensure that the freedoms associated with works in the analog world—including the freedoms to quote, recombine, and reprint—will not be sacrificed in the digital world by contractual restrictions. More generally, library negotiators should anticipate further new opportunities in the digital library and consult with prospective users and application developers who can advise them on the likely effects of contract terms.

Of course, the institutional partners should strive to safeguard their rights to explore and benefit from emerging applications that make use of the digitized materials. But these rights also benefit the business partner, as each of the areas in the list above may become a fruitful area for another form of public-private collaboration down the line. And libraries happy with the outcomes will likely refer more business to their partner.

**4. Sharing Content with Others**

The range of entities that have an interest in the assets and skill-sets developed in these partnerships is very broad. These include sister institutions in the fields of education and culture; commercial partners and potential partners; competitors for markets, resources, and attention; and the public.

Unfortunately, many of the agreements signed to date contain contract terms that prohibit pooling of digital files outside the originating institution, effectively leaving many of the opportunities solely in the hands of the commercial partner.

Interviewees in our project noted that current agreements stipulate significant restrictions on making all or part of the content available in aggregations—furthering in many ways the restrictions on use and reuse cited above. Many agreements include restrictions on wholesale downloading, commercial downloading, automated access, third-party redistribution, and open access to the work product of these partnerships.

On the most basic level, those engaged in noncommercial and noncompetitive activity should have the ability to work with the resulting content from a variety of these partnerships. When the commercial partner defines permitted users and permitted uses— and precludes combining the content with other content— the value of the effort is greatly reduced.

**Recommendation: Strive for the Freedom to Aggregate Content**

We recommend that all parties consider, much as they consider use and reuse, the range of potential stakeholders in the assets these deals create, and provide leeway in the agreements for future opportunities to cooperate with such stakeholders. These can include scenarios where noncommercial institutions collaborate with one another, where commercial companies work together, and even where hybrid groups cooperate to develop new products and services. The Open Content Alliance has been particularly concerned about safeguarding aggregation and content pooling, as noted in their statement of principles. A compromise is to allow the business partner to restrict distribution for a defined period, but after that period, the institution is free to share it with others.

**5. Circulation Records, Usage Data, and Privacy**

The protection of patron, usage, and circulation data is fundamental to library culture. In 1939, the American Library Association's Code of Ethics stated that "It is the librarian's obligation to treat as confidential any private information obtained through contact with library patrons."

Recent federal government investigations into library usage have highlighted the tensions between law enforcement and libraries over library usage and circulation records. Today, Article III of the ALA Code of Ethics notes: "We protect each library user's right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted," and the ALA's Policy on Confidentiality of Library Records states: "The Council of the American Library Association strongly recommends that the responsible officers of each library, cooperative system, and consortium in the United States[...] formally adopt a policy that specifically recognizes its circulation records and other records identifying the names of library users to be confidential."

All this stands in stark contrast to the practices of typical web services, which make broad use of patron, usage,
and "circulation" data as a matter of course. This usage data may be incredibly valuable to both libraries and their partners. Usage records and log files can be used to improve search services, collection management, and security. Improvements in search and personalization will benefit from access to usage data of various kinds. And yet this data also has potential for misuse and could put a chill on free inquiry by users who know that what they read may be monitored.

Though numerous federal and state laws govern the collection, use, and protection of patron information and other personally identifiable information, to date few if any of the agreements have explicitly addressed the collection, ownership, sharing, protection, and use of log file data.13

Recommendation: Include Usage Data in Negotiations

Data gathered by commercial partners about the usage of digital content should be provided to partner libraries. Partner libraries should ensure that personally identifiable information (and information that could be used to deduce personally identifiable information, such as IP addresses) should be secured against compromise and removed from log files when possible. Usage logs scrubbed of personally identifiable information will still be useful in supporting the development of new search tools.

6. Exclusivity, Duration, and Survivability

While these contracts are represented as non-exclusive, typically that non-exclusivity refers to the original materials. We hope that a library would not allow a private partner to dictate what the library can do with its original collections.

The business partner defends non-exclusivity by saying that the library is free to enter into another digitization partnership (or to redigitize the content itself). The fact of the matter is that, when taking into consideration the time and expense on the part of the library to participate in the first digitization partnership, it is very unlikely that the library will be able to repeat that investment when a second potential partner comes along.

Concerns about exclusivity should, in fact, be focused on the digital deliverables. This is the point of the library’s involvement in the undertaking. It is perhaps reasonable (though not inherently necessary) to allow the private partner a period of exclusivity regarding the digital files, in order that they can achieve their business objectives and recover some of their investment. But allowing them to restrict the use of those files indefinitely precludes the library from achieving its objective for entering into the partnership.

Just as deals that appear to be nonexclusive may in practice be exclusive, deals with fixed terms may, in fact, be far more lasting than they appear. Term lengths of the deals we reviewed range from five to thirty years. Yet while the term lengths are limited, in some of the Google contracts, at least, many provisions survive termination, notably those involving ownership and use restrictions. That is, the institution’s use rights to its digitized material may remain limited long after the commitments of its commercial partner expire. Are agreements with no fixed term on the limitations of use of the digital content really acceptable?

Most of these agreements auto-renew unless one of the parties opts out. Given that the Google commitments end at the end of the agreement and that the restrictions on use live into eternity, Google has little motivation to renew the agreement at the end of the contract term. Another aspect to watch is the flexibility of the agreement—what happens when a partner defaults? Do the survivability clauses terminate, for example, if Google ceases to offer access to the digitized content? Are there any situations in which a library could opt out of the agreement?

Those who defend having entered into agreements that are effectively exclusive and perpetual will often argue that the cost of scanning will decline and books will be rescanned again for quality, new uses, and other reasons. But as Paul Duguid has noted about the microfilm experience: "We got stuck for a remarkably long time with a retrograde technology poorly implemented because the start up (and opportunity) costs and barriers to entry, once the original microfilms and microforms were in place, made doing it all again prohibitive."14

Recommendation: Avoid Contract Exclusivity and Perpetuity

We recommend more careful study of the termination clauses in these agreements. Exclusivity on the right to manage and distribute the digital content should be limited to the term of the contract.

Safeguarding Rights for Future Use of Digital Content

Predicting the future of digital technology is a risky exercise. Transformations of media—from paper to microfilm, from radio to television, from print to the Internet—have always resulted in surprises. Just as it was assumed that television would be the end of radio, our guesses at the ultimate uses of digital materials are likely to look somewhat naïve a century hence. That said, it is possible to imagine users of the digitized collections from libraries, archives, and museums seeking to safeguard the ability to:
integrate digitized assets into teaching and learning materials;
freely use integrated digitized assets in text mining and linguistic analysis, such as exemplified at Columbia University's new Center for Digital Scholarship and Research;
freely share and integrate digitized materials with other researchers (as in the model of the Text Creation Partnership\textsuperscript{15});
publish and distribute digitized materials in a range of ways;
discover all digitized contents regardless of source or location.

Assessing the risks and rewards of these partnerships and the interests of our communities will allow us to develop materials that are accessible, sustainable, and interoperable, and that facilitate collaboration and experimentation for the public good. If limitations on this desired outcome are accepted, it should be with awareness of the compromises being made and their impact on potential uses. Practices that make future research and teaching applications difficult, or restrict communication within and between institutions, or perpetually limit an institution’s control over the digitized files developed from its collections will be deemed unacceptable by future generations, if not by ours.

Looking Ahead

Over the six-month term of our project, interviewees discussed how to evaluate digitization agreements against basic standards—of openness, of academic utility, of flexibility and control—in new taxonomies that others might find useful.

Several interviewees suggested rating the agreements by what they permit the community and the contributing institution to do, similar to the uses permitted by Creative Commons agreements.

One suggested a continuum of openness, beginning with the most open agreements, such as those of the Open Content Alliance (it’s interesting to note that the OCA, led by the Internet Archive, now has grown to include nearly 100 libraries, as well as commercial organizations such as Yahoo!, Xerox Corp., and Adobe Systems Inc.). In the middle are agreements like the one between NARA and iArchives, where the few restrictions are lifted at the end of a five-year term. Finally, at the other extreme are partnership deals that impose perpetually restricted access on the institution’s digitized materials, and those that limit what can be done with the originals, as well as the digital copies, like the Smithsonian/Showtime arrangement.

Another interviewee suggested that the community critically assess the impact each new agreement has on the overall academic and scholarly environment—in a sort of environmental impact statement for access and scholarship.

Whatever metrics might be used for appraising these rewards and risks, both the institutions and the companies will have to make sense and create value out of digitized content in a universe of hundreds of billions of pages of content where many of those pages are similar if not virtually identical.

As the stakeholders at the New York meeting explored their new role as producers and partners in the production of digitized content, they discussed the following, possibly contradictory, approaches to future negotiations.

First, they agreed that seeking digitization funding from commercial companies should not be addressed in isolation from the traditional fundraising strategies that libraries and others have pursued for decades.

Second, they agreed that these opportunities presented by Google, Microsoft, iArchives, the Open Content Alliance, and others really do represent unique departures and merit creative approaches to new business relationships. Should contributing institutions receive revenue shares, or even equity, as compensation? Are there other new business or legal arrangements that should be explored?

Several participants agreed that understanding the objectives—short- and long-term—of their commercial partners is important for succeeding in these relationships. They also reminded us that different types of institutional partners will have different mandates—government agencies like the National Archives, state-funded institutions like the University of Virginia, hybrid institutions like the Smithsonian, and private universities such as Princeton operate under different rules and aim toward different goals.

Participants desired a legal analysis of the publicly available contracts. Others encouraged the group to bring on business advisors to investigate new models of financing, perhaps even involving partnerships between commercial investors and philanthropic foundations.

Participants discussed reviewing the actual financial costs of these projects, and exploring whether the commercial partners are contributing equitably. Do they recognize that, while in some cases, they are investing millions of dollars, the content that the libraries bring to the table was preserved and cared for over time in processes that cost the institutions a far greater amount?

The point is to attempt to anticipate the future—the user of the future, the technologies and laws of the future, and the commercial partners of the future. After all, ten years ago, who would have anticipated Google?


13. A more complete discussion of this topic is in development by Jack Lerner and colleagues at the Samuelson Law, Technology, & Public Policy Clinic at UC Berkeley.


15. See <http://www.lib.umich.edu/tcp>.

Public/Private Mass Digitization Partnership Resources

This is a selection of materials about mass digitization, public/private partnerships and the agreements that govern them. It is maintained at <http://www.oclc.org/programs/ourwork/collectivecoll/harmonization/massdigresourcelist.htm>.

Google Agreements


Google Library Partnerships Commentary


NARA Partnerships and Related Commentary


Smithsonian Partnerships and Related Commentary


Other types of Digitization Arrangements


Policy


Related Commentary


Acknowledgments

The authors wish to thank OCLC’s Ricky Erway for her expert advice and commentary; Jen Mohan, who volunteered at Intelligent Television, for her expert analysis of contracts and support of our June 11, 2007, meeting; and numerous reviewers of this paper from the library and legal community.

Copyright © 2007 OCLC Online Computer Library Center

The U.S. Copyright Office recently published a report that addresses one of the most controversial areas of copyright law right now—the mass digitization of books. According to the agency, the report is intended to facilitate further discussions among the affected parties and the public regarding possible approaches to the issue, including voluntary initiatives, legislative options, or both. The nonprofit-commercial partnership concept is an established model. 24 See Helft, supra note 11. 9. B. The Nature of Existing Mass Digitization Projects This Analysis focuses on the purpose and objectives of digitization projects that involve copyrighted books, especially books that are available in the commercial marketplace. These objectives are likely to vary depending on the actor, and there may be more than one goal implicated in a particular project.Â be best suited to tailor collective licensing solutions for mass digitization to the evolving digital marketplace. A more dramatic approach would be to create a mandatory or statutory licensing. 2007. “Good Terms: Improving Commercial-Noncommercial Partnerships for Mass Digitization.” D-Lib Magazine 13 (11-12). 17 Thanks to Laura DeBonis, Jennifer Parson, and Jodi Healy at Google for reviewing the information presented in this section of the paper. Additional information about the Google Book Search is available at http://books.google.com/intl/en/googlebooks/about.html. 18 Thanks to Jay Girotto, Jessica Jobes, and Michel Cote at Microsoft for reviewing the information presented in this section of the paper. 19 Kirtas Technologies: http://www.kirtas-tech.com/. 20 Microsoft Live Search