As I started to think about digital special collections, I found myself pitching into a morass of unknown quantities and speculation. The dominance of electronic journals and aggregations of databases, followed by the news-grabbing mass digitization program at Google, has diverted attention from the fact that some publishers and libraries have been digitizing special collections for a couple of decades or more. For most of that time, an even greater number have been evangelizing the benefits of digital access to unique and rare materials. There was a time when I thought I had a clear vision of the future for sharing digitized versions of special collections with the world of scholarship and researchers far and wide.

Making high-quality images of special collections available on the Internet has opened up for archivists, curators, and librarians some dizzying possibilities. We now find our collections at the center of our libraries’ attention far more often than they were twenty years ago. We have a new set of choices in the realms of preservation, reformatting, and security. We are able to pursue high ideals for sharing a common cultural and historical inheritance by digitizing rare and unique materials for a worldwide audience. Now we can return the documentation of traditional cultures in digitized versions to the communities from which those collections originally came. We play new roles, with new partners, in placing “our” collections at the heart of the learning and educational experience. New forms of research develop, facilitated by the materials that we bring into the scholarly domain and the links and software developments that we create with our colleagues in information technology and the electronic industries.

The digitization of special collections has always been a complicated picture. Complicated by problems of scarce resources, and expectations that continually outrun the available technology and expertise. Complicated by politics, legal issues, and organizational boundaries. Complicated very often by the ways in which cultural and historical ownership interacts with the responsibilities and values involved in stewardship of the original materials in our care. Often, for example, the very communities we seek to serve, or ought to serve, are also those from which our special collections derive. Ownership issues in the digital environment are no simpler than those surrounding physical artifacts. Given the acceleration of mass digitiza-
tion, the continuing conflicts between publishers and libraries that show no sign of reaching resolution, and an exponential increase in the technical solutions that are on offer, visions of the future seem even more elusive.

We deploy in our community words like “discovery,” “exposure,” and “disclosure” (or, at a more detailed level of specificity, “preliminary record” or “collection level record”) as if one of these, despite the nuanced differences in their meaning, holds the key to the evolving policies that we must formulate and embrace. In a gathering of experts in this field, one ventures with trepidation to define the “big picture,” knowing that every chosen word potentially carries with it a host of different meanings. We live in an age when words are often carelessly thrown around or intentionally misinterpreted, resulting in giving and taking offence and generally massacring the meaning of everything. I do not wish to “misunderestimate” the potential for misunderstanding in the topic before us. But I do hope, to quote Mrs. Malaprop, a character in Richard Brinsley Sheridan’s play, The Rivals, written over two hundred years ago, that as I venture forward in this difficult and sometimes contested field, I can do so without attracting too many “aspersions on my parts of speech.”

What are Special Collections Today, and What Will They Be in the Future?

In 2007, the Association of Research Libraries (ARL) established a Special Collections Working Group, which I currently chair, made up of ARL library directors, heads of special collections, and several ARL leadership fellows from special collections libraries. Members of the working group seek to learn from and build on the large bodies of work that have already been done in this area, including that of the Rare Books and Manuscripts Section (RBMS) of the American Library Association, as well as the predecessor of our own group, the ARL Special Collections Task Force (2001–2006), which had already made an impact on the special collections agenda in ARL libraries.

The Working Group on Special Collections is charged with:

advising the Research, Teaching, and Learning Steering Committee [of ARL] on special collections issues on an ongoing basis. In this context, “special collections” is construed broadly to include distinctive material in all media and attendant library services.

The two issues that the Steering Committee identifies as the first priorities for the attention of the Working Group are to:

1. Identify opportunities and recommend actions for ARL and other organizations that will encourage concerted action and coordinated planning for collecting and exposing 19th- and 20th-century materials in all formats (rare books, archives and manuscripts, audio, video, and the like).
2. Identify criteria and strategies for collecting digital and other new media material that currently lack a recognized and responsible structure for stewardship.4

These two issues are closely linked since much valuable material in many different formats remains uncollected and at risk of being permanently lost. We need to develop coordinated strategies for identifying, collecting, preserving, and exposing these materials if we are to be successful.

I ought to make clear here, in case anyone doubts it, that the focus of the ARL group is not intended to suggest that materials predating the nineteenth century are unimportant or not at risk: it is simply a pragmatic attempt to give to the group a manageable task.

While international efforts are already underway to support the digitization of 19th- and 20th-century newspapers and books, even more are needed. Yet even before such work is possible, the culturally significant materials from these periods must first be identified and acquired, and both of these tasks are ultimately the responsibility of individual libraries and archival repositories. ARL can provide the leadership for encouraging collective activities, which include (but are not limited to) collection analysis (in other words, determining if and where gaps in the collection may occur); coordination of efforts; and encouraging the use of a “preliminary record” for identifying and making accessible collections that would otherwise remain hidden.

In addition to noting the top priorities in their charge, the Working Group also identified some general issues that should be addressed, including:

- identifying examples of how special collections contribute to innovative research, teaching, and learning;
- contributing to the work underway within ARL to develop qualitative and quantitative measures for the evaluation of special collections (which might include a target for surfacing hidden collections and mechanisms for tracking progress);

4. Ibid.
• contributing to and/or validating the work being done by the ACRL/RBMS Core Competencies Task Force to define the skills needed for work in special collections.

Periodically, the Research, Teaching, and Learning Steering Committee of ARL (which is the umbrella committee for this working group) may ask it for advice on other issues. For example, once the Steering Committee discusses a report from the ARL Task Force on the Future of Preservation in ARL Libraries, the Working Group may be asked to address the preservation strategies for special collections in both physical and electronic spheres.

In the context of this charge to the ARL Working Group on Special Collections, I would like to add some of my own thoughts on the subject.

We are all, I believe, obliged to take stock of the way we define “special collections” in the modern world, beyond merely asking questions about special collections in a digital environment. We should also consider how we have construed our responsibilities for these materials from centuries past. The leaders of ARL libraries, which contain some of the largest concentrations of special collections, are acutely conscious of this issue. It has been a common experience in the research community, often in response to new scholarship or because we ourselves have led scholars toward neglected material, to rediscover the intrinsic value of different sorts of materials, eventually leading to revelatory new insights.

To use one category of special collections as a metaphor, consider the case of printed ephemera. At one time, few libraries collected in this area beyond the fugitives and strays that may have been slipped between the pages of rare books, or those that turned up among the papers of statesmen, literary figures, or local dignitaries. It was left to the hobbyists—those maniacal amateur collectors of menu cards, bus tickets, beer mats, postcards, wine labels, playing cards, and so on—to insist on the importance of such materials. Faced with the insistent donor, we might provide an occasional home for such items, even while throwing up our hands at the challenges of preservation, storage, and the task of providing any kind of descriptive standards. Librarians and archivists still need to explain to the enthusiasts why we must be selective, but the professional community is now far more alert to the research value of this sort of material than we once were.

Among printed ephemera today there is still, I would say, a distinct hierarchy of materials. Anything that we might construe as relating to “arts of the book” or “print history” will generally find a place among the aristocracy of library ephem-

5. Ibid.
era. Playbills are theater history, of course; and posters, especially in such subject categories as politics or sport, constitute a class of their own. But I know, with some lingering contrition that dates from my days as director of special collections at the British Library, how difficult it can be to persuade even some of the greatest and most comprehensive research institutions of their kind that philatelic collections possess true research value. Many of us have our own personal favorite form of ephemera, but probably no two of us would define the word “ephemera” in quite the same way, even if we share a certain rueful recognition of some of the adjectives that Roget brackets with “ephemeral,” including “impermanent,” “transient,” “insubstantial,” “fragile,” “fleeting,” and “fugitive.”

In other words, the task of selecting collections worthy of preservation and incurring all the costs associated with their responsible stewardship has never been an easy one, even in the pre-electronic world. That task does not disappear in a digital context but is sharpened and changed in important ways, raising questions that never came up before. For example, how do we define a “born-digital special collection”? Does it have to do with the way we handle it, as well as the special skills needed to do so? Does this category include almost anything that is NOT an online journal or published e-book? What about second- or third-generation reproductions, such as digitized slides? Does e-mail correspondence fall into a particular category, or do we treat it as a records management problem that only archivists will know how to handle? E-mail also raises issues of literary creativity, legal questions, and the problem of “discoverability,” a quite different one from the innocent term “discovery,” and much more. A concerted effort is needed to define our terms in the electronic environment. Meanwhile, archivists and rare book and manuscripts librarians, including all of us who are charged with the care of multimedia collections in numerous formats, face a perplexing array of controversial issues that require difficult choices.

This article focuses primarily on the digitization of special collections, by which I mean rendering analog materials into electronic format. But, as archives increasingly arrive in hybrid form, we may also be challenged by collections that have been digitized in part already, though not necessarily in what we would consider to be complete or ideal versions. Consider, for example, published aggregations of rare books, or editions of newspapers that omit the advertisements. The activity of digitizing frequently coincides with making choices about how to treat materials that are already available in some kind of digital form. So it is important to clarify what we actually mean by the term “special collections.”

The Environment of Digitization

The Special Collections Working Group is preparing its recommendations for ARL.
The following major areas of concern will be included:

- Collecting carefully (with close attention to the total costs of caring for and making available a collection, be it in digital or some other format);
- Advocating against restrictions on access;
- Providing transparency, both in terms of provenance and the source of acquisition;
- Developing good practices in records management;
- Ensuring discoverability and access;
- Addressing the hidden collections problem; and
- Meeting the digital challenge.\(^6\)

The remainder of this article will focus on the environment in which the digital challenge presents itself, as well as some glimpses of the current scene, and finally a kind of peering, hopeful look into the future. Our collective thinking about digital special collections inevitably grows out of our experience in the physical, analog world and must be informed by it.

Before I move on, let me issue an invitation. The two national organizations that embrace most of the professional wisdom on special collections in research libraries, archives, and historical organizations in North America are the Rare Book and Manuscript Section of ALA (RBMS), and the Society of American Archivists (SAA). The Association of Research Libraries is essentially a meeting place for the directors of research libraries, advised by experts in the many fields that our organizations encompass. Although we are wonderfully supported in our working groups and committees by program officers, specialist fellows, and some co-opted professionals from the particular areas of work affected by each committee, we can only be truly effective if we have drawn deeply on the professional experience and insights of the community, and we can only be influential if we have their support. As soon as we release our draft report for circulation, it is essential that the Special Collections Working Group receive feedback from members of RBMS and SAA. Having received and incorporated that input and with the report accepted by the Board of ARL, the next stage will be a forum hosted by ARL in October 2009 to consider issues raised in the report, with a special focus on special collections in the digital environment.

The environments in which individual members of this community work differ greatly. There are huge disparities in resources, many differing missions, and any number of different collecting traditions. The following generalizations, however,

---

\(^6\) At press time, the report was not yet available, but will be posted and available soon at the ARL homepage.
apply to most digitization programs. Three infrastructural elements are fundamental: long-term preservation, provision for access, and the ability to repurpose digitized versions (for instance, in classroom teaching or for selective publication).

First, digital repositories of anything approaching adequate size or sustainability are still the exception rather than the rule. On the other hand, increasing numbers of universities are putting this infrastructure in place, at a cost of millions of dollars. Faculty members increasingly expect this infrastructure to be the norm. The presence of a trusted repository that can hold and sustain the digitized collections of the university’s libraries and museums as well as the vast array of databases and image-based research produced by faculty in all disciplines will come to be a defining quality of the leading research university. Where the technological know-how and the capital are not available in a single institution, partnerships and collaborations will most likely form to address these needs. State institutions—the classic example being the California Digital Library—can sometimes facilitate such relationships more readily than private universities.

It is important to note that smaller-scale independent scanning facilities now exist in most libraries and archives and that the majority of these institutions provide Web sites on which they can display images of some of their treasures. Only a few are totally without the capacity to provide scanning on demand. Indeed, it is through ad-hoc digital services of this sort that many special collections departments in university and college libraries first began their own digital libraries. Preservation issues form part of the context, and I will return to this question further along when I discuss the topic of mass digitization. The knowledge and experience of the special collections staff and their attendant metadata skills, as well as standards, inevitably affect decisions about digitization as well. Most ARL libraries, and large archives (such as those that derive from national organizations), possess at least some of the infrastructure needed in terms of equipment, capital investment, and staff.

Yet, for a large number of smaller organizations throughout the United States and around the world, this is not the case, and it seems to me that there are compelling reasons for collaborating and forming consortial arrangements. One of the strongest arguments for higher education to establish a robust, sustainable infrastructure for digital special collections (whether within single institutions or through consortial agreements) is the enormous increase in the use of primary resources for teaching and independent undergraduate research, which creates an inexhaustible appetite for digital versions of materials, either to supplement the original or to protect it from overuse. Equally compelling is the scholar’s need to search large quantities of rare material from a single location. Electronic serials and large
scholarly databases already provide this convenience, and there is an exponentially growing appetite for such services.

I do not want to neglect, in this brief survey of the environment and market for digital special collections, the broader societal benefits of making our resources available throughout the networked world. Digitization of collections that relate to particular communities, be they local or defined by ethnic or cultural origin, can provide social and educational benefits to users, as well as bringing political benefits to the originators. There is, for example, a continuing explosion of interest in local and family historical research. Political pressures to extend access to unique assets include the recent proposal by Senators Charles Grassley and Peter Welch to tax the wealthiest not-for-profits (including universities such as Harvard and Yale) by five percent if they do not spend a specified proportion of their endowment for the purpose of providing increased access to their resources. Providing access to special collections can help to meet that goal.

Countervailing pressures exist, however, including the thrust to protect intellectual property. Scrutiny of material available on the Web is intense and unpredictable; and making special collections available digitally carries with it the potential penalties of higher visibility. As just one example, Cornell University recently faced a lawsuit from an alumnus who objected to seeing a decades-old unfavorable story about himself on the Web in the digitized student newspaper. Digital special collections need to be packaged. They carry with them a danger of misinterpretation that cannot be mitigated on the Web through personal intervention as it might in a reading room or classroom. The huge opportunities that digitization presents to extend the accessibility and benefits of special collections undoubtedly carry with them potential drawbacks in the form of increased claims to the ownership of content, and a growing number of contested interpretations.

The Current Range and Scope Of Digitization in Research Libraries

If the description I have presented above is a recognizable representation of the environment in which we work, then what sorts of programs are flourishing within it? Each of us will have our favorite. There are so many local or subject-based enterprises that it is impossible even to characterize the landscape in a general way. Which are the projects that will turn out to be the foothills of mature mountain ranges, or the streams that become tributaries of some great Mississippi of digital content? (The Amazon metaphor, of course, is already spoken for.) Among my

---

own favorites are some of the digital initiatives currently taking place at Yale University designed to support teaching and learning. Numerous courses are now supported through interactive collaborations between faculty, librarians, and teaching and information technology specialists.

Among the proliferation of collaborative digitization programs that build on existing collections in museums, galleries, and archives, as well as libraries, my personal all-time favorite is Documenting the American South. This wonderful project, based at the University of North Carolina, was set up by Joseph Hewitt, who chaired the ARL Special Collections Task Force that preceded the ARL Special Collections Working Group. DocSouth, as it is often known, is a paradigm for the digitization of special collections within a broad theme. Anyone not familiar with it should take a look at its Web site. Do not neglect the letters and e-mails from users of the site, including descendants of slaves and slave owners as well as an overseas MA student and numerous scholars and community organizers, which vividly demonstrate why digitizing special collections, in an intelligent and programmed fashion, can be so worthwhile.

Mass digitization programs are generally projects involving multiple libraries: from the perspective of each individual library, the task is too great to be taken on without the significant corporate involvement and economies of scale that exceed the capacity of even the largest research library on its own. Perhaps the most well known are those that have been undertaken by Google, the Open Content Alliance, and the recent, relatively short-lived, Microsoft Live Book Search. Mass digitization of specific formats, such as newspapers, is also becoming an attractive option (which raises the question of whether or not newspapers should be considered “special collections”). One such project is building on the work begun by Early English Books Online (EEBO), an established digitization project basically composed of aggregations from numerous sources.

The impact of Google Book Search has been momentous, despite the culture wars that threatened at one time to break out between the Anglophone and Francophone digital empires or the fierce debates in both the scholarly and general press about the merits of digitizing large quantities of books. Google’s extraordinary coup, which started by incorporating five of the largest research libraries in the English-speaking world and now extends across the world, has changed the way libraries do their business. For a brief and digestible conspectus of many aspects of the Google impact, I commend the articles that appeared in the Journal of Library Administration in 2008 on the “Googlization of Libraries.”

---

Among the issues that particularly affect digital special collections are those that pertain to the quality of the output. Output is not always well presented, and can include folded-over pages and the occasional intrusion of an operator’s finger, not to mention problems with inconsistent and sometimes inadequate metadata. The quality of handling is also a source of anxiety. Google’s machines have not been demonstrated widely, and there are concerns arising as much from ignorance as from empirical results, about the way in which books are treated at the point of scanning. On the other hand, the libraries currently working with Google are not reporting significant numbers of damaged books returning from the scanning operation, and Google has been working hard to improve the quality of its product.

More intractable is the question of copyright. When the legal dust has settled (and that may not be soon) there will still remain many questions. One of my own great concerns is about the impact of presenting just 10 percent of books that are in copyright on the whole educational enterprise. It seems to me that facilitating the use of small snippets of a book out of its overall context does violence to the principle of scholarly argument. Several leaders in the world of research libraries (including Charles Henry of the Council on Library and Information Resources, or CLIR) have expressed serious doubts about Google’s business methods, including the extremely high degree of confidentiality, locked in by nondisclosure agreements with each participating library, that have made it impossible for the library community to discuss adequately, still less to influence, the process whereby millions of their volumes appear on the Internet. Given the tight contractual restrictions on reusing the digital files, it is a serious concern and limits the universality that surely was one of the great attractions of this enterprise in the first place.

One undeniable benefit that has come from the Google mass digitization program is its influence. The Open Content Alliance, for instance, while not enjoying the wealthy resources of a corporate behemoth, has made available a large quantity of digitized material. Its open source principle makes it possible for libraries and archives to share their digitized content whenever it is free of copyright restrictions, and the potential benefits to research from this collective approach are breathtaking. Microsoft’s rival approach, with the Live Book Search project into which several libraries (including the British Library, as well as Yale, Columbia, and Cornell universities) entered, may have been short-lived, but it will leave a legacy of expectations. Even as we at Yale wind down our Microsoft operation, we value the careful handling, and closer engagement with the library, that characterized Microsoft’s approach.

I will not list all of the now numerous separate mass digitization enterprises that have sprung up alongside Google, Microsoft, and the Open Content Alliance. But
there are a few more consequences that should be mentioned as an important part of the environment. Scholars now have an appetite for increased quantities of digitized material. This in turn has an impact on the policies of individual universities in helping to build the support we need, and making the case for capital investment in a robust infrastructure. We are now able to pay close attention to collection development issues. It was a revelation at my own library to learn from a preliminary analysis of OCLC’s Worldcat that we have something like one million volumes that are not held by any of the original Google Five libraries, and they too, presumably, must have huge quantities of unique holdings. Mass digitization has undoubtedly spurred OCLC Programs and Research to develop its collection analysis tool, a great step forward that will inform libraries’ local decisions about collection development, as well as lay the groundwork for future digital partnerships. Here, perhaps, is one possible route to the Holy Grail of collaborative collection building—if not in the physical world, then in the digital one. And here for sure is a tool in the hands of all of us who wish to increase the availability of our special collections in the digital environment.

The Future

What priorities derive from our responsibility as the stewards of multiple inheritances?

First of all, one critically important recommendation of the ARL Special Collections Working Group is that there should be no digitization without metadata. Discovery, that primary function of research in special collections, is not possible without description and guidance. The “Hidden Collections” agenda, as defined at a conference organized by ARL at the Library of Congress in 2003,10 has moved into the center of the stage, and the importance of surfacing special collections is essentially what that agenda is about. A new program of grants from the Council on Library and Information Resources, supported by the Andrew W. Mellon Foundation, explicitly exists to promote cataloging and description, not digitization for its own sake. With a budget of $20 million to be spent over five years, it will make a significant contribution to the fundamental task of exposing and disclosing material for which our organizations on their own lack the resources to place in the public domain.11 No digitization will take place without the metadata. The future, in this respect, may look like the past but, we hope, with more resources to support the work we need to do. And those resources will necessarily be spread thin. CLIR is looking for projects that will provide models of good practice for the lean and effective description of collections.

that until now have been hidden altogether from scholarly inquiry. Access begets access: let us get the stuff out there, and then work on it some more when, and if, it becomes clear that further descriptive work is needed.

That still leaves us with any number of outstanding issues as we attempt to shape the future of digital special collections. For example, what is the future for born digital collections, and how do we influence that? And what exactly do we mean by “born digital collections”? At this stage, we really only know for certain that there are specific skills we have to develop and issues we need to address, including:

- Digital curation
- Developing new relationships with our users: how do we connect with them?
- Working with our users to describe collections
- Mirroring traditional functions in OAIS (Open Archival Information Systems) models
- Coping with volume and scale
- Identifying partners
- Being prepared to wait for new technology to help achieve more satisfactory curation

Finally, as we contemplate the future, we need to think about the contest over standards and metadata, platforms, decisions about commercial versus in-house development, and the battles that are in progress and yet to come over adequate funding for infrastructure. As we survey the emerging outlines of this still clouded future, we can at least tell ourselves that special collections librarians and archivists are well prepared for such battles and uncertainties. Our work will always be a work in progress. We are not going to achieve perfection. But, as we strive together to invent and reinvent our future and the future of our collections, we can at least aspire, as Mrs. Malaprop might have hoped, to be models of the “very pineapple of politeness.”