

Reflections About Collections: With the Help of Peter Cook and Dudley Moore

Ann Okerson (Yale University Library)

ann.okerson@yale.edu

Charleston Advisor

Published in Volume 7, Number 1, July 2005

Ed. Note: The following paper was delivered as the opening keynote presentation at the Seventh Fiesole Collection Development Retreat, in Melbourne, Australia, April 28–30, 2005.

Introduction

My generation of Canadians loved the famous skits by Peter Cook, Dudley Moore, Alan Bennett, and Jonathan Miller called *Beyond the Fringe*. Clueless, or wacky, or dilettante, or demented, or totally out-of-it Brits sat around and pondered the mysteries of the universe from their unique life perspectives, perhaps a coal mine or a mountaintop. One of my favorites in *BTF* is “Aftermyth of War.” Here, heroic English folk contemplate life during the Great Second World War, consoling themselves throughout the bombings by having “a nice cup of hot tea.” “Squiffy,” the wife says to the yokel towards the very end of the sketch, “it’s the end of an era,” and this time they console themselves by having “a nice cup of boiling hot water.”

We are going to take the next short while to reflect on whether we are reaching the end of an era, through the voice of a librarian as often as not in the coal mine of collections. But please understand that, as I look at the current and future academic research library scene, particularly with relation to collections development, the right or best strategies for moving forward with a long-term collections strategy aren’t particularly clear to me.

My primary message today is that none of us believes our libraries are self-sufficient, and they are not. Yet, in spite of evidence to the contrary, we too often act as if we are still very much autonomous.

Building Library Collections: The End of an Era?

Libraries go back centuries, but the ambitious library, focused on and open to a broad community of users, is, for the most part, a very modern concept. The antecedents of the Yale Library—troves owned and operated by scholars of extraordinary learning and brilliance—were brought home to me recently. We have just finished cataloging and converting bibliographical records for one of these unique old enclaves, the Yale Babylonian Collection, numbering some 10,000 items, i.e., as large as the whole of Yale Library was until late in the nineteenth century. Throughout this project, a joint venture between our library and the distinguished faculty curator of the YBC, we heard stories about the “old” days when faculty and graduate students would drop their own work and go on “book hunts” through the warren of YBC rooms, in order to find a book for a visiting scholar who needed a particular item. Collections in that storybook world were built with much learning and much intuition; we now build teams of specialist bibliographers, scour the world to fill our gaps, and scoop up new publications.

In the 1960s, huge amounts of money came into our hands in the U.S., for both new and old universities and libraries, all growing faster than staff could manage. So Richard Abel and Co. (Oregon) hired a cadre of subject selectors to do a piece of academic book selection that was being repeated daily in libraries all over the country. His company then supplied its selected books on “approval” to libraries, which could choose to accept or return the books. It was a brilliant idea, “outsourcing” before the word was a commonplace. Now “outsourcing” is increasingly common, as academic booksellers, and not just in western countries, provide additional services such as online catalog copy or “shelf-ready” books upon delivery. Were our libraries to

reclaim the selection, acquisitions, and cataloging work that has been entrusted to our commercial colleagues, our staff costs would surely at least double.

Next, in this postmodern library age, large aggregators of electronic books are entering the marketplace: for example, netlibrary, ebrary, and ebooks.com are among the newcomer generalists, while specialists such as Safari, Knovel, and Books24x7 offer narrower ranges of collecting expertise. As e-book consolidations proliferate, library selection becomes even more fully outsourced and we in libraries have less flexibility to reject titles. The Yale library now “holds” well over a half million volumes of e-books purchased and licensed, none of which were chosen as single titles (and a number of which replicate our print collections). All that is before the promise of the Google Print and other mass digitization projects.

Let me emphasize that all the developments described here reduce local collections and selection activity, and—here’s the hook that gets us—greatly expand the accessibility of materials. But, at the same time, wherever more and more selection work is done outside of libraries and/or collections are licensed by us, content available through research libraries, at least in North America, is becoming more similar each year.

The journals collecting enterprise is in some ways different from book collecting, but in other ways it is now similar. Selecting a single journal biases future acquisition: once having selected a journal, the librarian’s default inclination has been to continue receiving the subscription indefinitely. Thus, until the 1990s, selection of journals—by contrast to monographs—already felt “efficient” and was done almost exclusively title by title, with the business support being taken on by subscription agencies (who, unlike book vendors, do not select for libraries).

That said, the zeal for seeing the benefits of electronic publishing was nowhere so intense in the 1990s as in the area of scholarly, and particularly STM journals, where change was rapid. I was among the earlier librarians involved in consortial negotiations for what we now call “the Big Deal” (licensing all or virtually all of a publisher’s titles in one financial arrangement), initially with Academic Press, in conversations that began in a corridor at the Frankfurt Book Fair in 1995. Through the late 1990s, we begged publishers for such scaled-up arrangements and rejoiced at the greater comprehensiveness of titles for a small premium in dollars. Now the pendulum has swung, and librarians more often speak of the Big Deal as the devil’s spawn—while at the same time, we are impatient with publishers of single titles who have not found the cash or technological savvy to integrate their product with the rest of the world’s e-journals.

Whatever one may think of approval plans, or of the rapid advancement of online technologies, or of e-book collections, or of e-journal aggregations and big deals, they have surely affected libraries’ collections development in at least two profound ways:

First: The Reader Rules. A world in which more readers have more places and times wherein they can access information, whether through libraries or online, through licensed materials or free sites, has made this an age of the reader in remarkable ways. Whatever else happens, and whatever we try to do, the reader today has unprecedented access and control over his or her information supply and needs—and thus over his or her demands and expectations of libraries. This inevitably is moving libraries from being collection-driven to being reader-driven.

Second: Where Doesn’t Matter. Technological advance has reinforced a long-term trend by expanding interlibrary loan and document delivery services. It is no longer as important to have exactly the right content in one’s building. For example, Yale is a founding member of a consortium of seven large book-lending and borrowing research libraries participating in a project called BorrowDirect, a patron-initiated process for selecting a desired book from any library in the consortium and placing an online request for its physical delivery. By 2003 we had achieved a fill rate of over 80 percent, up from 43 percent; a response time down to four days on average (about as

long as it takes for that Amazon book that you ordered to arrive); and transaction costs reduced by some 60 percent. This has been an immense success, and, as you know, many other library consortia are providing similar services.

Organizing and Preserving Collections: Down in the Coal Mines

Let us leave for a moment the topic of autonomy in book selection and collections development and turn to library self-sufficiency in organization of materials. In the early 1970s, for example, it was possible to see simultaneously in the great libraries of Dublin, Ireland, three generations of technology, all now swept into obsolescence: old bound book catalogs at the National Library, a home-built and handmade card catalog at University College, and vendor-supplied card cataloging at Trinity College. Each newer technology repeated the pattern of mechanization and outsourcing described above for books and journals. The Library of Congress began selling catalog cards in the U.S. in the early 1900s and OCLC and RLIN began to develop their services in the 1960s and 70s. If Yale Library had to do original cataloging for all the materials we buy these days, we would surely be spending \$5 to \$10 million per year on this activity.

Sometimes research libraries' collecting policies have us purchase books we cannot catalog immediately, against the happy day when we will find resources to do so. For example, we are now recruiting for an 18-month position, which requires subject knowledge of South Slavic or Hungarian literature or culture, or knowledge of one or more additional Finno-Ugrian or Slavic languages. The University has provided some funds for acquisitions, and we now have a backlog of some 4,000 books in these languages, titles that do not appear in the large cataloging utilities or databases. Nor can these booksellers supply us with catalog records. If we are successful, the 18 months of money we have will take care of less than half of the backlog, but we may not be able to find someone for this work.

That is, the biggest research libraries are still unable to polish off various catalog backlogs in unusual languages, or in special collections and archives, let alone the new media (particularly electronic) coming on board in tsunami-like quantities. There is simply too much "stuff," much of it only in electronic form delivered over the Web. Increasingly, we don't even speak of cataloging anymore—we call it "assigning metadata," some of which is extracted by computers. Of course, much more such automation needs to be done, though it is easy to argue for the "better" and "higher quality" record that a skilled professional cataloger can create.

And here's an important budget conundrum for collections development officers such as I. As catalog records are offered by book or e-vendors, those access costs are more and more commonly added to or bundled with the costs of the content. The lines between collections and access, already a little blurry, overlap and intermingle all the more in budgetary negotiations, whether for catalog records or servers or software. These days in virtually every major research library, we engage in a great deal of back and forth across the collections-access line, openly or surreptitiously. Which budget pays for access? To what extent do the actual collections themselves matter and where does one draw the boundaries?

A quick word about preservation: As the growth and variety, formats, and media of library materials have accelerated, the challenge of preserving content has grown, outstripping completely libraries' financial and human resources available for the job. There is far too much content for us even to dream of comprehensive preservation initiatives. Our Library of Congress is taking a national lead through its NDIIPP project, and other national libraries and groups are working hard on digital preservation technologies and models. LOCKSS, for example, offers a caching solution, a controversial solution but one libraries can implement locally, even if for, so far, a limited range materials, to keep them safe for a time. The Andrew W. Mellon Foundation sponsored Ithaka group will unroll its Portico e-journal preservation service this June. So far, both LOCKSS and Portico are just for journals. Other components of the problem remain largely and frighteningly unresolved.

Thus we rely increasingly on publishers as the authoritative sources for e-journals, reference works, and e-book collections. We see control slipping away, with no local or outsource model in hand yet to give us confidence. And, of course, Google, with its widely-touted digitization initiative, seems poised to join this outside-the-library group on whom we will have to rely, but is reticent about its plans and so can do little to assuage our anxieties.

Enter the Rabbit

We have quickly reviewed decades of library outsourcing and automation—did we really think outsourcing had been invented only a few short years ago? If you are a Monty Python fan, you've seen *The Holy Grail* (movie) and perhaps even the new Broadway play, *Spamalot*, with the enormous Trojan rabbit being pulled to the French castle by King Arthur's knights. We might imagine the rabbit to be full of chocolate delights, or we might fear that heavily armed knights will leap from it to destroy all that we hold dear. But, then, perhaps the *Spamalot* rabbit offers none of these predictables and thus there may be no easy resolution to our story.

Present: The Trends of Today

Let us bring our attention now to the present, the real present. To see the present, it is necessary to discover the trends, curves, movement all around us. Let me identify five key components of today's library collections landscape.

First. Making an Electronic Marketplace

When the e-information age arrived in force (mid to late 1990s), we had few precedents for arrangements between e-information providers and libraries, let alone consortia of libraries contracting for content on behalf of all their campuses and readers. Everything needed to be spelled out. In the give and take, back and forth of negotiations, libraries and providers began to learn how talk to each other productively, to come to some accommodation (often in such an agreement, everyone is a little unhappy, which can be the hallmark of a decent arrangement). At this time, license agreements have come to describe the network of business relationships among various constituencies of the electronic community. In fact, the license became the means by which it was possible to create an electronic marketplace. It is even conceivable that license terms will become so standard for various categories of content (such as e-journals) that we will not need them much at all; all that will be left will a pricing agreement—for, as we all know, price continues to be the greatest area of dissent and disagreement between libraries and their suppliers.

Second. Economic Churn and Uncertainty

The cost of production and the prices particularly for electronic publication are in seemingly unending churn. The conversion of print journals to e-journals is rapidly maturing. And while e-books have yet to find their way into the commonplace in academic libraries, the recent huge initiative of Google Print to make e-books indispensable may accomplish that transition at supposedly a cost of about \$10 per scanned volume. Meanwhile, the rapid development of aggregations of electronic content has made possible the distribution of far more information, and far greater usage, per dollar for electronic resources than for print. And, aggregations of libraries are pushing down price increases, though not necessarily prices, somewhat. The big challenge today comes as these consortia are trying to dismember the e-journal Big Deals from larger publishers, unwilling to give up much of the access that has been gained in signing up for these deals in the late 1990s. For example, because of budget constraints on the part of a growing number of our members, our consortium (NERL, with 26 ARL-sized libraries and 42 smaller affiliated institutions) has begun downsizing talks with a handful of big publishers whose three-year package deals are set for renewal next year or the

year after. This will be a defining moment. One has offered us an interesting opening proposal. If we wish to cancel, say, 5 percent of our package price, library members participating in that particular arrangement would need to cancel their immediate access to e-journals that received 5 percent of the use – we could choose whether this would be one 5 percent title or multiple titles whose use totaled 5 percent. Gives one a headache, doesn't it? I won't even mention the churn and allure of Open Access, which has earned its own program segment at this conference!

Third. Rise of the E-Aggregators: Good News or Bad?

We librarians are quick to notice when large journal publishers buy other journal publishers or bid to produce under contract journals that were previously individually published or published by small producers. Less noticed but of equally high impact has been the rapid and amazing growth of content offered by third party consolidators or aggregators of e-journals and databases. Here are some not-for-profit aggregator numbers:

The ALPSP Learned Journals Collection was launched in 2003 with 247 journal titles from its member publishers. It now stands at 433 for 2005 and growing. This is a partnership of largely not-for-profit publishers with Swets.

BioOne began in 1999 with 40+ journals; it now delivers 81 titles, with aims for over 200. This is a largely not-for-profit partnership including societies and libraries.

Project Muse, launched in 1995 by the Johns Hopkins University Press and Library, currently electronically publishes over 250 titles, mostly in humanities and social sciences and mostly from university presses; Muse regularly adds new titles to its collection.

Some of the largest or growing for-profit aggregators include EBSCO, Factiva, Lexis-Nexis, and ProQuest—with thousands of titles each, many of which fall in and out of the databases with terrifying unpredictability. Note that just a few years ago Kluwer (Ovid-SilverPlatter) was the largest single source producer of indexing and abstracting services, with about 150 databases. Quietly, it has been overtaken by the electronic publishing side of EBSCO Information Services, now offering 150 e-databases, of which it owns about half (and offers them pretty much exclusively), as well as 5,500 active full-text, peer-reviewed journals. Recall my earlier comments about consolidation and outsourcing of library collections decisions!

Fourth. Institutional Repositories

An institutional repository is a beautiful concept and a necessary service. The idea began, I believe, with the philosophy that research institutions should stop giving away the information they produce and care for it in perpetuity at home, while enabling access for free for the world. Initially intended to solve one problem, IRs are now being constructed to solve another much more significant one: to capture many or all of an institution's productivity and activities: university archives, departmental records, faculty and student works such as databases, editorial projects, geospatial and other data, images, courses and course materials, and much more. Every institution needs to make such provisions for its creations. Doing this is decidedly not cheap. As we undertake it, I confess that taking care of material already being published and archived elsewhere may not be the highest priority.

Fifth. Google Print

By virtue of the spectacular quantity of digitization (at least by today's standards) Google has established a leadership position in the e-marketplace, which brings revenue, recognition, and control—all without doing very much yet! Smaller initiatives may be obscured for now (those of Brewster Kahle and Carnegie Mellon

University's "Million Books Project" and Amazon's "Inside the Book," to say nothing of thousands of smaller digitizing projects in many libraries). By aiming for huge scale, Google Print becomes a powerful symbol or metaphor for all digitizing that is happening and will happen. Today, in 2005, the future of Google is a great unknown, and a potentially destabilizing one for existing players in the content "space." In the news yesterday (April 27, 2005), we all read that 19 European national libraries are petitioning their governments to establish Google-sized projects, in order to assert their own leadership and cultural patrimonies. Meanwhile, I asked one of my colleagues in one of the Google institutions what they were doing differently with regard to collections development. The answer was: nothing really. We are proceeding with our same book purchasing programs, licensing electronic resources, digitizing special collections. Google plans to index some 20 million books in the OCLC database, and yet our library life goes on as it always has. How can this be?

Four Future Strategies and Paths to Future Collections Success

I can gaze into no crystal ball. To think of the future is to develop principles and strategies to embrace opportunities that come to us – big, scary, and exciting, in real time, opportunities to deliver more and better content and support to our readers.

First. Think Systematically

We need to learn to think of the totality of the research information universe in which we live and have our being, to think with keen and accurate quantitative sense, and to identify our role in that universe. Successful services depend on it. Along those lines, I was most impressed by a new study reported at the Coalition for Networked Information's (CNI) meeting earlier this month by Brian Lavoie (OCLC research) and Roger Schonfeld (Ithaka). Using OCLC's WorldCat, they have made some careful first order counts of the size of the "monograph universe" in libraries, the distribution of copies of individual titles, and the rates of growth over time. The data are in some cases stunning. For example, they indicate that half the books ever published in the world—titles, not just individual copies, half the titles – have appeared since 1977. They estimate 26 million individual monograph titles (i.e., nonjournals, nonserials, nongovernment documents) in the 20,000+ OCLC member libraries. They observe that about 24 percent of the individual titles are held by more than 10 libraries, and only 5 percent of the individual titles are held by as many as 100 libraries in OCLC's database. The 9 million titles held in only one OCLC library risk becoming orphans if we "digitize from the top," so to speak. "Rareness is common," they observe with a wry sense of provocation. (One other sobering number: about 83 percent of all the books ever published appeared first after 1923—in other words, within the presumptive range of the American copyright protection period. The public domain is much smaller than you think.)

What does this mean for our collections activities of the future? Should librarians build collections? Become publishers? Why?

Second. Take Advantage of Systemwide Savings

To be sure, we libraries have organized ourselves, and not badly, in order to deliver information to our users. Our consortia license resources for the benefit of many libraries; we enter into cooperative arrangements to share books (Borrow Direct); we contribute to and derive our catalog records from one another. Each library belongs to numerous, almost uncountable groups and partnerships that are helpful in some way. That said, I am pretty sure we have not gone far enough, at least not in an efficient way. Here are some of the specific questions I have about my own environment, for example:

Why did we and each of our peer institutions go through its own RFP process and contractual arrangements for separate Integrated Library System packages, a process that easily took us two to

three years, a great deal of human resource, and some “lost-opportunity” time, in addition to money—so that now we must develop tools—often they have large limitations—across those ILS systems in order to do cost-effective resource sharing?

Could we have developed our remote shelving facility with other libraries in our region for more efficient access and de-duping of materials?

Could we develop our future collections by spending money more effectively, i.e., focusing on our strengths and developing those, while looking to partners for materials in which they excel?

Why does each of our libraries load catalog records for commonly held electronic resources into its own online catalogs?

Does each institution need to build completely from scratch, at considerable effort, its own institutional repository?

Third. Build and Contribute Through Specialization

Do we want each to invest in collections that are increasingly similar to those of our peer institutions? What if we thought instead that we might want to get past our traditional horizontal library collections (lots of everything for everybody, defined by the diversity of people at the university) and chose instead a vertical model (information in depth for a given scholarly community to use worldwide)? Perhaps we could not so effectively pursue such a strategy for print materials, but is there not great potential for specialization of our electronic effort? The keen-eyed and often contrarian observer and consultant Joseph Esposito advocates exactly such a shift, reasoning that (1) mass digitizing cannot and will not be comprehensive; and that (2) to the extent that the Stanford/Michigan, etc., books are rare or idiosyncratic (rather than those held by hundreds of libraries), they will present a false picture of the cultural record, especially when machines are increasingly heavily used to search this record. So, should we “digitize what Google left out?” Or is that an ineffective and inefficient way to proceed? Might we in academic libraries better utilize our resources through careful and planned specialization, so that the cultural record is not just Google’s record but the record of many specific domains?

Fourth. Effectively Develop Partnerships and Alliances

“We shall all hang together,” said Benjamin Franklin, “or assuredly we shall all hang separately.” I have tried in this paper to signal how our past and to some extent current stand-alone practices lead us to continue to imagine our libraries as more self-contained than not. I’ve shown at least briefly, and without any attempt to be exhaustive, some need to make common cause. We librarians still do not as easily recognize our huge interdependence on other libraries, consortia, shared collections to satisfy our own readers—let alone our interdependency on publishers and suppliers. We have a long way to go to learn how to work together. Just why that is could be the topic of another essay.

Here the challenge is first to see, then to make ourselves act: see what we are now doing, match that with our goals (of building collections and serving readers), develop ways in which we might most effectively build collaborations and cooperation, then act to build these wisely and well, or even to risk beginning unwisely (though not stupidly!) and thereby learn important lessons. Such thinking will be most effective if we imagine cooperation on a grander scale than most of what we do together now (of course, there are some exceptions). Is there leadership that can bring us together effectively to re-imagine ourselves in this way? What would such leadership look like? Are we at risk from always thinking too small? If the issues I have outlined do not persuade you that we think too small, then let me throw a few more onrushing asteroids into the field of vision.

The explosion of content provision in a world of blogs, vlogs, podcasts, and the like. There is serious material there that needs to come inside the fence of intelligently managed and accessible information. These are the collections of the future, along with the about-to-be mass digitized library collections. What have we done to make them available?

Two words: China and India. The universe of people who consume serious academic information is exploding right now, today. If we think that a few hundred independently funded, academic libraries working separately and collaborating on small- to medium-scale projects are ready for this, we are, I have to say, nuts.

Let me remind us all again of the Lavoie/Schonfeld numbers and the breathtaking fact that half of our printed books are under 30 years old. We realize, on that scale and even with a limited sense of geometry, that the new half life of printed information will soon be 20 years and less. In other words, there is reason to think that the collections and content we will need to manage will more than double within our lifetimes and probably double twice or three times in the lifetimes of the very young. Growth of that scale cannot be managed by business as usual.

My friends, the journey of a thousand miles begins with a single step. Which one shall we take next?

The Charleston Advisor © Copyright 2005 Ann Okerson