

September 2006

Collection Development & SURVIVAL in the "Mostly Digital" Library

Michael Stoller

New York University Libraries, michael.stoller@nyu.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Stoller, Michael (2006) "Collection Development & SURVIVAL in the "Mostly Digital" Library," *Against the Grain*: Vol. 18: Iss. 4, Article 8.

DOI: <https://doi.org/10.7771/2380-176X.4915>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Collection Development & SURVIVAL in the "Mostly Digital" Library



by **Dr. Michael Stoller** (Director, Collections & Research Services, New York University Libraries, 70 Washington Square South, New York NY 10012; Phone: 212-998-2566)
<michael.stoller@nyu.edu>

As researchers have seen their access to remote materials increase, local collections have lost much of their significance. We have a unique opportunity to share, an opportunity that we have, on the whole, lamentably ignored. The moment has not entirely passed, and we should still take advantage of this unique chance. But, as paper collections decline in importance and digital resources move to the center, we'll face a new challenge that will dwarf the old resistance to collaboration that came from institutional pride.¹

The "Mostly Digital" Library

We are facing the dawn of the "mostly digital" library, a world in which the vast majority of primary and secondary materials on which scholarship rests are in digital form. It's "mostly digital," because some components seem unlikely candidates for digitization any time soon. The vast archival holdings of our special collections, millions of linear feet of correspondence, manuscripts and realia, while they will certainly be the objects of selective conversion and the source of wonderful digital exhibits, don't at present suggest a viable business model for mass digitization. And area studies may lag behind, though Latin America, the Middle East and even Africa may surprise us and turn to digital publishing more quickly than we imagine. But the paper vestiges of the journal world will surely become digital before the end of the decade, and, at some time very soon, that last great bastion of paper — monographic publishing —

will find viable business plan(s) for digital delivery or face extinction in a world of 500-copy print runs and remaindered back lists. Even textbooks and mass market publishing may soon convert, as students rebel against the outlandish cost of the former and a viable reading platform emerges for the latter. **Google, Yahoo!, Microsoft** and a host of entrepreneurs we've yet to imagine will digitize our legacy collections. The large, paper-based collections that have defined research libraries for centuries may not vanish, but they'll be little used, and the scholar of the next generation, like many scholars already, will spend most of her time at a computer.

Does this mean that the library, if not already "overwhelmingly a virtual destination," as **Jerry Campbell** has said, soon will be?² Absolutely. Does this mean, as many writers seem to think, that collection development is fading away like **Marx's** state, to survive only as license management. Absolutely not. The latter does not follow naturally from the former for a simple but often curiously ignored reason: these digital collections will not be free. Libraries will still need to decide how and where to spend their money, and, if we're not all to bankrupt the institutions we serve, we'll need to be able to shape those collections carefully, as we always have, and to exercise far more control over what we buy than the world of digital publishing has offered us thus far.

One thing we can say with some certainty:

in the mostly digital world, the local collection will not have the meaning that it had in the paper age. It is a privilege to browse the shelves of a great research library. I think it always will be. The scholar exiled at a small college with a tiny library lived a harsher research life than colleagues at a great research university with a rich library. But today a large part of that great research collection isn't on the browsable shelves. It's miles away in a remote storage facility, browsable only in the catalog. Just as accessible, however, are the collections of dozens of other research libraries in a world where interlibrary loan is fast and getting faster and remote collections are nearly as accessible to the small college professor as to his colleagues at the research university.

But there is a problem with this democratic vision of our new library world: it applies only to the paper collections, and, in the mostly digital library, those paper collections won't matter very much. What will matter is a scholar's ability to search the literary holdings of *Early English Books Online*, the journals of *Science Direct* or the government documents of the digital *U.S. Serial Set*. Scholars will want to search the full text libraries being built by **Google, Microsoft** and others, libraries that will surely not come free. But, unlike paper books, these digital resources, as the library world is structured right now, are not an interlibrary loan away. They're licensed to the institutions that buy them, and their use is almost entirely limited to affiliates of those institutions. In the "mostly digital" library, as it's evolving right now, that scholar at a small, rural college is back in the same researcher's exile that typified such places in the 1970s.

Reinventing Collection Development


This is exactly why collection development needs to be at the center of our thoughts, not a forgotten remnant of the days when we bought our collections one book at a time. As we wrestle with how to build this mostly digital future, we can't abandon collection development. Instead, we must reinvent it. It's not a local enterprise anymore. It's not about matching the book to the faculty member. It's about making sure that we have a truly national collection, that scholarship and the raw materials that are its sources — materials that will be mostly digital — are accessible to researchers everywhere.

So, let's talk a little about the reality of the mostly digital library. First, as anyone who's watched it grow in the last decade knows, it won't be free. Open access will be a part of the picture, at least in the sciences. But even that won't be free. If true open access models emerge, and scholars pay publication fees or

continued on page 30

If It's Not Digital from page 26

useful and to integrate it with their way-of-working. Do not waste time explaining. We "just" need to supply what they need quickly, profes-

sionally, and on their desktop in friendly and, if possible, interesting ways. This will require resources (obtained from stopping some of the things we do now that do not support the above), investments in flexible but standards-based and Web-based IT systems, and a re-aligned workforce. 

References

1. **Bergstrom, Ted** and **Preston McAfee**. Journal Cost-Effectiveness [Website] "search engine to find internationally-published journals and rank them by price per article or citation." See: www.journalprices.com.
2. **Osorio, Nestor L.** (1988) Approval Plan Materials and Direct Orders: Differences and Similarities. In **Grayson, Lawrence P.** and **Biedenbach, Joseph M.**, Eds. Proceedings American Society for Engineering Education, 1988 Annual Conference: Beginning a New Era in Engineering Education 4, pp. 1729-1735, Portland, Oregon (USA).
3. **LaBonte, Kristen B.** (2005) Citation Analysis: A Method for Collection Development for a Rapidly Developing Field. *Issues in Science & Technology Librarianship* no. 43 (Summer 2005) p. 1 at: <http://www.isrl.org/05-summer/refereed.html>.
4. **Levy, Steven** and **Brad Stone**. (2006) The new wisdom of the Web. *Newsweek*, April 3, 2006, p. 47 - 50, 52 - 53.
5. **Wilson and Hayward**. (1999) Winning through chaos—part 2. *Credit Control*, 20 (9/10): 35-39).
6. **Wittenberg, Kate**. (2006) "Beyond Google: What next for publishing." *Chronicle of Higher Education*; Section: *The Chronicle Review*. Volume 52, Issue 41, Page B20.

universities make their research available on institutional repositories, the money will have to come from somewhere, and only the most foolish university administrator won't look to the library's material budget. And we've already seen would-be open access ventures, like the *Stanford Encyclopedia of Philosophy*, come to library's for "contributions." Freed of the old-fashioned subscription fee, we're asked instead to be venture capitalists. But fees or contributions, it won't be free.

Nor should it be. For all the wonders of the wiki, of "open editing" and social tagging, the formal process of publishing is not something the academic world is likely to abandon. Peer review and the professional editor have been fundamental to the quality assurance structure that lies at the heart of modern scholarship. They will take new shapes, and one hopes they'll find ways to be economically viable without crushing our budgets. But it's still going to cost money to turn the raw output of a scholar's work into something others will want to read and be able to trust.

As for the legacy collections, **Google** and **Microsoft** are businesses, just like **Elsevier** and **ProQuest**. It is a wonderful thing that they've decided to bring the resources of the commercial world to digitizing our collections. Using the budgets of the not-for-profit world, it would take generations to do what commercial ventures can complete in a few years. But they'll not do it as a public service. They'll let scholars look at their indexing. But the content will come at a price. We won't want patrons wading through pop-up ads, and it would be abdication of our responsibility to ask them to get out their charge cards. So we're going to have to license content, buying back our collections in the new format just as we've bought back all the books former microform vendors have sold to us.

If the current licensing landscape remains, we'll lose much of the ground for collaboration that we've gained in recent years. We'll have our consortia, of course, but that's not collaborative collection development. It's just collective bargaining. Consortia are simply a way for us to gang up on our vendors, to present a united buying club in hopes of securing lower costs. The real ground for collaboration has been created by new efficiencies in interlibrary loan and by the changing environment generated by offsite storage.

Libraries Again As Isolated Islands

But the mostly digital age presents new obstacles to collaboration. The new obstacles are legal, written into licenses and enforced by the courts. They can take us back into a world where each collection is an island, and some islands will definitely be richer than others.

In this environment, collection development will be more critical than ever. I don't think it will be the same. We'll need to reinvent it. On the one hand, we will still have to build local collections that match local needs. On the other hand, and far more importantly, we will need to

build a thoughtful, national collection, collaborating to make access as ubiquitous as it now is in the paper world.

On the traditional level, collection development will still involve some title-by-title selection. Aggregated purchasing, the "big deal," has shown itself to make little sense. It's a publisher-driven model, a great way to pawn off the low end of their lists — whether journals or monographs — packaged with the high-demand titles. We've started to rebel, and I think that will continue.

There is a good reason for aggregation, of course, one the vendors like to flaunt — data mining — the ability to search across vast aggregations of content and retrieve information without respect to the specific journal or book in which an article or chapter found its original home. And data mining is certainly the future. But data mining isn't more valuable, if the aggregation includes a publisher's poor journals along with the best. One digs for diamonds where they're likely to be found. Anyone who says he's offering you a deal by letting you dig for diamonds in Sweden along with South Africa is just trying to swindle you. My neuroscientists are not likely to find their data mining enriched, if I purchase an aggregation that includes agricultural journals along with the bio-science titles. We'll need to take back control of what we decide to buy.

That's true for books as well as for journals — perhaps more so. We can't afford subject specialists spending endless hours evaluating the latest monographs. That's not a good use of their time in today's library. But we're still going to have to devote substantial time to making sure our patrons have access to the books they need, whether those books are in paper or digital form. Buying great aggregations of monographs from a digital vendor makes no more sense than loading up my library's shelves with paper books my patrons don't want. If digital monograph vendors try to take us down the evil path of aggregation, we'll need to "just say no." Aggregated searching, mining monographs for information, will be a blessing in the digital world. But we'll want to make sure patrons are mining in rich veins.

Far more important, from my perspective, is the need to develop our collections nationally. It is we who need to aggregate. It is we who need to build a data mine accessible to scholars at libraries large and small, rich and poor. We must translate the democratic landscape of interlibrary loan we've built into an environment that works in a mostly digital world.

That doesn't mean signing licenses that allow us to print out an article from a vendor's database and ship it or fax it to a user at a smaller library. That's what we do today. We allow some remote, unaffiliated researcher to glimpse a small piece of one of these enormous collections — a journal article, perhaps even an entire digital book. That's a sort of literal replication of our old interlibrary loan world — letting someone borrow a small piece of our collection.

But that will be a meaningless gesture in the mostly digital future. As I've noted, much of

scholarship will be about data mining, searching across large collections of material. The interlibrary loan model doesn't apply to that vision of scholarship. Interlibrary loan is about known-item searching, a scholar looking for a specific book, a particular journal article or document. Using that model, researchers at large institutions will be trolling in the rich waters of vast electronic collections, using the powerful search engines our vendors are building for us. Researchers at small institutions won't have access to those powerful search tools. While their colleagues at **Harvard** and **Michigan** are searching full-text databases, researchers at small schools will be relegated to searching the far more limited universe of bibliographic utilities like **WorldCat**, hunting for known items. Being granted the small privilege of borrowing a single 16th-century title from *Early English Books Online* is nothing like searching the full contents of that collection. If the scholar at a poor institution wants to experience that glorious power, she'll need to travel to a great research library and beg the privilege of sitting in its reading room and using its collections on-site, just as her predecessors did with our paper collections decades ago. I'm sure we don't want to go back to that future.

I wish I could say exactly how we can go about this national collection development. It was hard enough breaking down the walls of pride that isolated our great collections in the old days. Indeed, we never really did. They just disappeared, when patrons discovered they could mine other library collections with quick, UPS-driven interlibrary loan. Breaking down the legal walls that vendors and their lawyers have erected around our digital collections will be far more difficult. We've fought long and hard for the few rights to control our own collections that typify the newest of our contracts. It is an infinitely greater challenge to break down that all-powerful **shibboleth**, the "full-time equivalent," that has governed licenses and tethered collections to those who happen to be enrolled in or employed by our parent institutions. As long as that one small phrase governs our acquisitions, the mostly digital landscape will be a world of haves and have-nots.

A New Sort of User Status

The answer might lie in an extension of the rights we give to visitors in our libraries — to the person willing to come to use our resources onsite — the license term is "walk-ins." It's a peculiar idea, that in this digital age, when resources can be accessed from anywhere, we have still clung to this notion that a guest of the library must actually walk in the door. Couldn't we create a new sort of guest — a "virtual walk-in?" We could build a clause into our licenses that allows us to grant a certain number of temporary, guests access passwords to a resource each year. Such access might or might not be broader for public institutions, as it is for walk-in access. But all of us need to take responsibility for building access to a national collection, and the privates have a role to play along with their more accessible, public peers.

We must break our digital collections free of the restrictions that have made them so dif-

continued on page 32

Is Access to Government Information Getting Better or Worse?

by Katherine Holvoet (Head, Government Documents, University of Utah) <kate.holvoet@utah.edu>

The simple answer is yes, access to government information is getting better and worse. Electronic indexing is improving access to material, and obscuring it. The **Government Printing Office** is attempting some heroic measures to improve access to government information, but doesn't have the budget to control the entire process. Commercial vendors are creating fabulous value-added government information products that are unattainable for many libraries. The upcoming wave of retirements among government documents librarians may create a significant loss of subject expertise, but will open opportunities for newer librarians. In short, there is cause for optimism and concern.

As government information moves online, it is sometimes easier to find material, such as tax forms, and sometimes harder. More agencies are creating databases for information such as technical reports, patents, and research publications. Search engines and Web crawlers will index the search page for these databases, but there are settings that must be present to allow a Web crawler to index the database itself, and some government databases are not set to be indexed by Web crawlers. This effectively places that information in a black hole. The average searcher would have to know exactly which database contains the information for which they are looking.

Currently, government documents librarians are the best finding aids for government information. Comprehensive computer indexing is one of the greatest advantages the digital format has to offer, and is not available

for U.S. documents. Mending this problem is technologically simple. Search engines could index the databases, but setting policies is politically difficult. The **Government Printing Office** does not have the authority to force executive branch offices such as the **Department of Homeland Security** or the **Department of Energy** to allow Web crawlers to index their databases.

For example, if a search is run in **Google** for "patent number 5,003,456," the result is one return, "Circuit for providing fast output current control in a current mode switching power supply." The resulting hit is from *freepatents online.com*, not from the **Patent and Trademark Office (PTO)**. The commercial site includes advertising, and may leave the researcher wondering if the result is in fact an official patent. There are over six million patents available online through the **PTO's** Web search, and none of them will show up in a basic **Google** search unless they are put online by someone other than the **Patent and Trademark Office**.

Perhaps the most talked about issue today revolves around preserving long-term access to born digital documents, and digitizing and preserving the historical government record. Who is responsible for archiving government information? Should that trust go to a government agency such as the **GPO**, or to library consortia such as the **Association for Research Libraries (ARL)**? What role should commercial publishers who offer terrific but expensive products play?



The **Government Printing Office** serves the pleasure of **Congress**, and executive branch departments serve the President, leaving open the possibility that inconvenient information could be purged from the public record. The significant change in access to online government information in the wake of 9/11 should alert librarians and citizens to the negative consequences of single institution provision of government information.¹ For example, the **EPA** has removed **Risk Management Plans** for facilities such as water treatment plants and dams, from their Website even though the necessary details a terrorist could use for planning purposes were never put online.² Now citizens no longer have ready access to pertinent safety information about the public works installations in their communities.

Cooperatives such as **ARL** may not have the finances or the organizational structure to effectively coordinate and implement such a massive preservation task. Attempting to copy and preserve born digital documents in the absence of a central database of government information would be maddening. Agencies no longer rely only on the **GPO** to publish their materials — now they can simply put them on the Web. Any born digital preservation project would need to continuously crawl all government Websites for new documents, and then save copies to servers. The task of identifying government information online, capturing a digital copy, storing it, and indexing it for easy retrieval may be beyond any organization's capabilities. The closest example to such a project is the **Internet Archive**,³ and the indexing for that is not very user friendly; it works best when one already knows the URL.

In regards to digitizing the historical record, consider the **US Congressional Serial Set** (the official record of **Congress**) which comprises over 12 million pages. Without centralized control of digitization and metadata cataloging, it would be very difficult to ensure every page met project standards. The best way to ensure centralized control would be for **ARL** libraries to all contribute money to pay to have the works digitized by a commercial vendor, and to have metadata cataloging created. Not all **ARL** libraries are willing to pay to create a digital **Serial Set**, and not all libraries would be able to pay even if they were interested. Another option would be to distribute the burden for digitizing the **Serial Set** among **ARL** members, along with standards and guidelines. Not all libraries have the expertise to digitize documents, or access to the necessary material.

What about commercial publishers? Could vendors fill the preservation need by digitizing government information for resale? The online **US Congressional Serial Set** by **Readex** is a thing of beauty, with tremendous added value in the extensive indexing. Quotes for this prod-

continued on page 36

Collection Development ... from page 30

ferent from our paper holdings. How curious that we so willingly gave away that control. Of course, we weren't trained to read contracts in the early days of digital resources, and we were doubtless blinded by the simple need to restrain the outrageously escalating costs. But now it is time to step back and ask ourselves whether we're truly willing to give

up the extraordinarily democratic infrastructure we've so recently built for access to our paper collections and to return to a time when a researcher's institutional affiliation defined his access to library resources.

Digital technology has made it possible for researchers to access our collections without regard to *where* they are — in the library, at home or in the office — we should not have to add the caveat that it now depends on *who* they are. 🍀

Endnotes

1. Many of the issues treated here grow out of a study I did recently of collection development patterns in research libraries during the last decade, particularly concerns about the divide between large and small, rich and poor. See my recent article: "A Decade of ARL Collection Development: A Look at the Data," *Collection Building*, 2006 no. 2, pp. 45-51.
2. Jerry D. Campbell, "Changing a Cultural Icon: The Academic Library as a Virtual Destination," *Educause Review*, January-February 2006, p. 20.