Interlending trending: a look ahead from atop the data pile

Dennis Massie
OCLC, San Mateo, California, USA

Abstract
Purpose – This paper seeks to explore five forces likely to significantly affect interlending operations in the near term: the transition from print to electronic resources; management of legacy print collections; mass digitization projects; competition from other information providers; and copyright.
Design/methodology/approach – The author uses data from authoritative sources to illustrate the effects these forces are having and will continue to have on libraries and ILL operations.
Findings – The author predicts that most libraries will be slow to divest themselves of print monographs on a large scale; libraries will continue to build new offsite storage facilities but put more thought into their contents; increased discoverability of digitized texts and greater copyright restrictions will drive users to print; librarians will make gray areas of copyright law work for them instead of against them; publishers, librarians, authors, lawyers, and scholars will find a responsible and fair solution to providing digital access to “orphan” works; and ILL will persist as a core operation for nearly all libraries.
Originality/value – This paper provides a unique look at forces that are shaping the future of global ILL activities.

Keywords Interlibrary loan, Library trends, Change management, Libraries, Resource sharing

Paper type Viewpoint

Introduction
In 2001, I was fortunate enough to attend the 7th IFLA Interlending and Document Supply (ILDS) conference in Ljubljana where there was a lot of talk about whether ILL had any future. Leaf through the conference program and you will find presentations such as “When resource sharing used to work” and “E-books, the demise of ILL,” and – my personal favorite – a panel discussion about publishers selling articles directly to end users called “Two’s company, three’s a crowd: who needs a librarian?”[1]. There was so much negative talk that I was tempted to change the title of my own presentation to “ILL: let’s call the whole thing off” which I intended to sing to the tune of that old “you say meta-day-ta, I say meta-dah-ta...”

You say meta-day-ta, I say meta-dah-ta...
You say final user, I say end user...
Digitize! Copyright! Publishers! Compromise!
Let’s call the whole thing off...

A decade later, ILL activity among Association of Research Libraries members is up 20 percent, from 3.8 million filled requests in 1999 to 4.5 million in 2009[2]. Filled requests on the OCLC ILL system are up 17 percent, from 8.7 million in 2001 to 10.2 million in 2010[3]. Obviously, the idea of libraries sharing collections with each other is still very much alive.

The current issue and full text archive of this journal is available at www.emeraldinsight.com/0264-1615.htm

Five forces, five years
I have identified a list of five forces that I believe will affect the course of resource sharing operations during the next five years. While the list is certainly not comprehensive, these forces at least lend themselves quite nicely to numerical analysis. By examining hard data, I will forecast how those forces will affect our behavior as resource sharing professionals. How will we be going about our business five years from now, and what will be the nature of that business? The forces under examination are:
• P to E transition (print to electronic).
• Managing legacy print collections.
• Implications of mass digitization projects.
• Competition.
• Copyright.

P to E transition (print to electronic)
What does the data tell us about the ongoing transition in our libraries – and indeed in our society – from mostly print to mostly electronic?

For one thing, it tells us that the transition has already happened with periodicals, scholarly and otherwise. We have got them scanned, backed up, archived, and stored away in dark caves, some never to be seen by human eyes again. Finally, librarians are feeling that it is safe to divest ourselves of...
the print counterparts to all these electronic journals. There are 7,160,028 articles available in JSTOR as of February 20, 2012[4]. In addition, there are now half a million pre-1923 articles available even to nonsubscribers (JSTOR, 2011). In most fields, even the humanities, scholars are going digital because of convenience and lower costs to libraries.

But it would be premature to assume that the transition to electronic journals is complete. Three years ago, Constance Malpas attempted to identify all peer reviewed scholarly journals currently published and came up with around 27,000 titles[5]. Moreover, an astonishing 36 percent, or 10,000 titles, mostly in the humanities, are available only in print.

These are not oddball special interest magazines like Weekend Cannibal. They are peer reviewed scholarly journals. These findings suggest that not everything will be digitized, not even all scholarly research will be digitized. Perhaps as much as one third of the scholarly literature will remain available only in print (and who are you going to call when you need an article from one of those print-only titles for your next research project? ILL, that’s who).

There is much more to be said about the transition from print to electronic. For example, the percentage of the library budget that Association of Research Libraries members spent on licensed resources more than doubled in a seven-year period, from 25 percent in 2002 to 56 percent in 2009 (Kyrillidou and Morris, 2011).

But it is not just libraries that are transitioning from P to E. If you go to the library of the University of California, Berkeley, to see if you can check out an electronic journal, you will be sent to a survey called, “Are you a subscriber?” If you answer “no,” you must order the journal via ILL. This is an important beginning.

What are the implications of all these numbers for resource sharing librarians? It is obvious that the stock we rely upon to service our patrons is changing. Library users are not limited to the paper publications that were available in the past. Fewer libraries will continue to operate as if they have a mandate to preserve print resources. Libraries that will continue to see preservation of print materials as part of their mission will need to self-identify sometime in the near future. In fact, this has already started to happen with efforts like the Western Regional Storage Trust (WEST) Project(6) where shared print archives are linking up to form a network and making their preservation commitments widely known. Libraries cannot all go E and forget about P at the same time, or essential research materials will be lost forever.

Managing legacy print collections

The second of the five forces influencing ILL is the management of legacy print collections. There are already nearly a billion print volumes in North American college and university libraries – wow! – and we are adding another 25 million each year (Payne, 2007), many of which are going directly into storage. That is quite a legacy we are leaving, especially since so many institutions are storing the same things.

Lizanne Payne conducted a study of storage facilities in North America for OCLC Research and found that about 70 million print volumes are stored in 68 offsite storage facilities across the US. Most of those facilities are full and many organizations are considering building new storage facilities or adding pods onto existing facilities. The cost of doing so is between $3 and $4 per volume, a very serious investment when you consider that such facilities are usually built to hold about 1.5 million volumes (Payne, 2007).

One might think it would be cheaper to leave those print volumes on campus, but Paul Courant determined that it costs $4.26 per year to keep a printed volume on the shelf vs $0.86 per year to store it in high-density offsite storage (Courant and Nielsen, 2010).

Another study found the price of keeping a monograph in a library over the course of that book’s lifetime is a cool 718 percent of the original purchase price (Lawrence et al., 2001). Other formats are not nearly so expensive to keep as monographs, but the price tag is still hefty.

What are the implications of all these numbers for resource sharing librarians? It is obvious that the stock we rely upon to fill ILL requests is an expensive pet to keep around. Filling most ILL requests is fairly easy at the moment because you can always find somebody that has what your patron needs on the shelf. But will that still be the case when administrators start looking for less costly alternatives to storing all these print volumes?

Not only are print collections huge and expensive, gobbling up resources desperately needed for other things, but they are barely even used. Study after study shows this to be true. A ten-year study of the ILL and circulation activity of 89 OhioLink libraries found that a mere 6.5 percent of the aggregate OhioLink collection accounted for 80 percent of the use (OhioLINK et al., 2011). Since the system averaged four copies of each and every volume, there was plenty of “nothing” to “not circulate” (O’Neill, 2009).

The University of California system and others with a long history of using offsite storage show that 1 to 2 percent of stored collections are paged each year (Payne, 2007). We have no idea what titles are in shared storage because, with a few exceptions, the records for those items are loaded into WorldCat under the owning library’s symbol, not the storage facility’s. Because most materials in high-density shelving are arranged by size, there’s about a 0 percent chance any of that
material can ever be weeded. So all that print that is widely duplicated and not being used is most likely there to stay. If all that stuff in storage is likely to stay there because of the difficulty in weeding material shelved by size, maybe we should start thinking of such collections as de facto archives. It may be worth finding out what is stored in those archives and seeing if all the storage facilities across the system could be leveraged as a collective asset to all libraries. Of such modest thoughts are mighty research projects born.

One such project became known as the Cloud Library Project (Malpas, 2011). It explored the idea that a shared print storage facility, such as the 8.5 million volume ReCap (jointly run by Columbia University, New York Public Library, and Princeton University) might provide the basis for a third-party library such as New York University (NYU) to consider divesting itself of significant portions of its print collection, assuming the third party library could buy subscription access to what is stored in ReCap. For this to work, the collections stored in ReCap must sufficiently mirror NYU’s print collection.

We also compared the print ReCap collections and NYU’s print collections with the HathiTrust corpus, the idea being that HathiTrust would provide preservation assurances for the materials, while ReCap would provide physical access to users. We expected to find substantial overlap among the NYU, ReCap, and HathiTrust holdings. After all, everyone collects and stores the same things. Right?

Well, not exactly. While there was a lot of overlap between NYU’s print volumes and HathiTrust, only 10 percent of those titles were also held in ReCap. The availability of print back-up was considered essential to support any divestment of print collections by the third party library. The good news is that by including the print holdings of not just ReCap but also of all campus libraries of all institutions that own ReCap, overlap with NYU’s print holdings represented in HathiTrust rose to a remarkable 90 percent (Malpas, 2011).

What are we to make of this? It is simple. We had been hoping that a library like NYU could divest itself of print collections by partnering with a single print storage facility that mirrored its collection and would provide its users access to the materials formerly owned by NYU. That did not pan out. But by expanding the replacement collection to include the full collections – not just stored stuff but everything – owned by three top-tier ARL libraries – duplication with NYU approached 100 percent. The next questions to be explored were as exciting as they were obvious:

- What if a network of print archives existed?
- With a business model in place so libraries could subscribe to just-in-case access to print materials?
- How many supplier archives would be needed?
- Could other libraries divest themselves of low-use print?

And, of course, there is a final question of particular interest to resource sharing librarians:

- What effect would this model have on ILL?

We added the collections of five major storage facilities to the mix. The Center for Research Libraries had the least amount of overlap with HathiTrust – less than 5 percent – but this is not surprising since CRL collects the kind of low-use ephemeral material that would only be found in a place like CRL. ReCap, with its 8.5 million volumes, yielded an overlap with HathiTrust that approaches 20 percent. The combined holdings of the University of California’s Southern and Northern Regional Library Facilities duplicates 45 percent of HathiTrust’s holdings. And Library of Congress duplicates almost 60 percent of the digitized material represented in HathiTrust. Combined, all five storage facilities could meet 75 percent of the print access need for items preserved digitally in HathiTrust (Malpas, 2011).

So, to return one last time to our original supposition, having a third party library establish a partnership with a single print storage facility would in most cases not provide adequate coverage to allow that third party library to winnow its own low-use print titles. But by adding just a few more storage facilities, the print duplication of the massive HathiTrust digital corpus grew to nearly 75 percent. Clearly, a network of linked shared print facilities would alter the equation for the better. But important questions remain. Would such a network really enable other libraries to discard print? And is there a business model that would make sense for both consumers and suppliers that would allow this to happen? Finally, what effect would outsourcing a significant portion of your low-use print collection have on your ILL traffic? Your own library might have little or no print collection, so you would be borrowing from networked storage facilities frequently.

Implications of mass digitization projects

Mass digitization projects such as HathiTrust constitute the third of our five forces because of their implications for the way library collections are built, managed and shared.

One thing you can safely say about the HathiTrust is: “Wow, it’s big!” According to the HathiTrust website, over 9.9 million volumes, or 5 million book titles, have been digitized already, including almost 3.5 billion pages[7]. HathiTrust also includes a quarter of a million serial titles, and it’s getting even bigger. With more partners joining, the rate of growth is skyrocketing. HathiTrust grew to 6 million volumes in just twenty months. Current projections are that it will equal the size of Harvard University collections (16 million volumes) by 2013, and surpass the Library of Congress (30 million volumes) by 2020 (Malpas, 2011).

Aside from its sheer size, one of the most interesting and important aspects of HathiTrust is the copyright status of the digitized items. It includes over 2.5 million public domain volumes, many of which happen to be US government documents. Overall, 27 percent of its collection is in the public domain and now freely available. Within HathiTrust, nearly three-quarters or 7 million titles are covered by copyright, searchable by many, but viewable in their entirety by very few[8]. Estimates of the number of orphan works – works published in years you would think would be covered by copyright protection, but where the rights holders or the copyright status of the item is difficult to determine – run as low as 2 million and as high as 5 million titles. A sample done by the University of Michigan copyright office found that 45 percent of 100,000 titles checked were orphan works (MLLibrary News, 2011). That projects to nearly 4.5 million potential orphan works in the HathiTrust.

The University of Michigan and a number of other HathiTrust partners (University of California, University of Wisconsin, Indiana University, and Cornell University) have established a rigorous method for checking the copyright status of suspected orphan works, and plan to release such works for full access to their users as titles are cleared. The first batch of 27 orphan works by 27 French, Russian, and American authors is about to be released. A quarter of a million researchers will have full access (Young, 2011).

For daring to push the orphan works envelope, several author’s associations have filed suit against these institutions
(Authors Guild, 2011a). Rights holders want to put a stop to this. One author said the HathiTrust takedown policy for presumed orphan works is like requiring a homeowner to notify the cat burglar ahead of time that he does not wish for his home to be ransacked. An author's guild blogger tracked down the agent for one orphan work's author. It turns out the writer had written a book upon which an Elvis Presley movie had been based, and he has just signed a contract with a publisher to release an e-version of yet another of his out-of-print works (Authors Guild, 2011b). Obviously, this is going to get ugly.

If HathiTrust prevails and all the orphan works eventually are made available to researchers at the schools involved, that means nearly 7 million full-text volumes would become available to a quarter of a million students. Of course, that is not the same as releasing all these works to all Internet users, but from a librarian’s perspective it is a step in the right direction.

The news gets better. There are nearly 5 million records in WorldCat for items represented in HathiTrust, with new records being added every day, and OCLC has just reached an agreement with HathiTrust to provide full-text indexing of the entire digitized corpus within WorldCat (OCLC, 2011b). Researchers will soon enjoy the benefit of an almost unimaginably powerful new tool.

I believe this indexing will drive more researchers to use the print versions of works, after they find what they need via HathiTrust or WorldCat but cannot view the material. A couple of studies, one at Michigan and another at Columbia University, have shown no evidence that being part of the digitized corpus results in greater use of its analog form. Helen Look examined University of Michigan students’ use of the top 500 titles from HathiTrust and, although students racked up nearly a million page views within these 500 works, only 2.2 percent of them circulated that year. Nearly 40 percent had never circulated (Look, 2010). However, students in the study had full access to the digitized versions, so why would they need print? It is when folks can discover text but not access it that we will see them driven to print.

**Competition**

The fourth of our five forces is competition, by which I mean competition for the attention share of information seekers. Competition can be bad if users bypass libraries in their information gathering and rely on inferior sources, which can have a negative impact on library funding. However, competition might also benefit scholarship, as more sources of information are brought together in one place. Some competition threatens to put us out of work. Some helps us do our jobs better.

I have chosen three competitors for profiling. One was named in *Perceptions of Libraries* as the point where 84 percent of researchers start their quest for information. Not a single person surveyed reported starting their search on a library website (OCLC, 2010). Our competitor? The internet, of course.

This next is an example of what can be a good, even helpful competitor. If an entity such as OAIster adds 25 million records for open access digital resources to the place where I search for books and journals, that entity is my friend[9]. The same goes for ArchiveGrid, which adds a million records to the world’s biggest bibliographic database[10]. This competitor would be aggregators.

Our third competitor puts up some really big numbers each year, such as 8 billion – that’s dollars of revenue from US scholarly publishing. Two billion of that comes from subscriptions sold overseas. No wonder this competitor – publishers, obviously – does not want us to have the right to supply copies of articles to libraries in other countries (Association of American Publishers, n.d.). My feeling is that if publishers spent as much time providing convenient, affordable end-user access to their resources as they do trying to make ILL lending from e-journals completely onerous, they could double their take and we would be happy for our patrons who were being served so well. But do not hold your breath. We will be printing and rescanning e-journal articles for a while more, I am afraid.

**Copyright**

The last of our five forces is the really big train hurtling down the track: copyright.

As we learned from that Cloud Library Project described above, each of the top-tier Association of Research Libraries members owns about 3.5 million print volumes and there is typically a 30 percent overlap between that ARL library and HathiTrust (Malpas, 2011).

For example, over 700,000 titles were held by both NYU (print) and HathiTrust (bytes). If copyright permitted and NYU could provide full access to all 700,000 titles, NYU could divest itself of the corresponding print volumes and save 55,000 assignable square feet of space and as much as $3 million in cost avoidance for storing those volumes (Malpas, 2011). The space saved would be more than twice the size of a typical learning commons. If every library could do that, libraries and scholarship would be transformed. ILL might finally go out of business.

But that is not likely to happen anytime soon. The Google Book settlement – as imperfect as it was, still our best chance to make that massive digitized corpus widely available – is pretty much dead. New copyright legislation may come along and ease the problem, or help librarians and publishers find some common ground. And the Chicago Cubs, who last won the World Series just four years before the Titanic sank, could become world champions again, too.

Librarians can push the envelope and take any privileges not expressly forbidden by copyright law until they are made to stop, what my colleagues and I call “running until tackled.” The University of Michigan’s orphan works gambit is an excellent example of this. Brewster Kahle and the Internet Archive amassing their own print repository of books may be another, as one suspects that part of their strategy is to posit some relationship between what they can do with their digital copy if they also happen to own a print copy (Wohlser, 2011). I certainly hope so.

**Predictions and conclusions**

Now it is time to make a few modest predictions about the way these five forces will actually affect the way we do resource sharing in the next five years.

First, libraries will not rush into unloading print. We have already seen how perfect circumstances had to be before libraries jettisoned print back files of JSTOR titles. The rest will be harder and take longer. Some libraries in extreme need are already withdrawing print but others will not do so as long as there is any good reason not to. Print will stay, accessible for a while more, I am afraid.

“Managed scarcity” will happen within the next five years only on a small scale. We simply do not have enough experience to reduce collections to the minimal number of copies needed to support the entire system. It is too uncertain
how increased demand for fewer copies will affect the condition of the physical item. It is clear that if we throw stuff away and then want it back, it is going to be just too darned bad.

Groups like the Western Regional Storage Trust will do a great job of sharing preservation commitments with others. Various libraries will assume responsibility for different parts of the collective collection. What will continue to elude us is a business model that would allow consumers to ditch print and buy subscription access to stored materials. I think that is a shame and a golden opportunity lost, or at least deferred.

I am convinced that having full-text indexing of HathiTrust in WorldCat, right next to that “find it in a library” button, is going to drive users to seek print copies of works in the digitized corpus. It is certainly how I behave, and I am not all that peculiar.

Institutions will continue to build print repositories, shared and otherwise, but will be extremely careful about what they place there. You will not see duplicates or JSTOR titles. More thought will be given to subject matter and past patterns of use. Filling these facilities will become quite a deliberate process.

Librarians will let go of old fears and boundaries and push for as much access as the law will allow, and make gray areas work for them instead of against them.

I have a (perhaps irrational) belief that “orphan works” is an area where publishers, librarians, authors, lawyers and scholars can come together and find a sensible solution that is fair to all parties while promoting research and creation of new knowledge. This seems doable, and the potential impact could be massive.

My last prediction is that the Cubs will finally win the World Series – certainly before ILL as we know it ever has reason to go out of business. But that could still lay a long way into the future.

**ILL word cloud**

Finally, I made a word cloud out of the titles and abstracts of papers presented at the 12th IFLA Interlending and Document Supply Conference (Figure 1). It contains ideas and issues discussed in the conference presentations. No matter how closely I look, I cannot find a discouraging word about ILL. No “demise.” No “stop.” I found the word “barriers” in the upper right, but just above it are the words “allow lending across.” I love the fact that “sharing” is almost as big as “libraries.” If I have one regret, it is that “user” is not a little bigger. But it is pretty big. Certainly bigger than “barriers.”

We have gotten very good at what we do. The ideas expressed throughout this conference prove that we are all striving to be better still. Our libraries and our users need that from us. Five years from now, they are still going to need us. The data pile says as much. And I am going to go out on a limb here and guarantee it.

**Notes**

Interlending trending: a look ahead from atop the data pile

Dennis Massie


5 Interview with Constance Malpas, 7 September 2011. For background information see www.oclc.org/research/activities/policy/default.htm

6 See Western Regional Storage Trust website for more information on the WEST Project, www.cdlib.org/west

7 Data obtained from HathiTrust website at: www.hathitrust.org (accessed 8 December 2011).

8 Data obtained from HathiTrust website at: www.hathitrust.org (accessed 8 December 2011).

9 For additional information on the OAIster database, see the website at: www.oclc.org/oai/stem/

10 Additional information on ArchiveGrid is available at: www.oclc.org/research/activities/archivegrid/default.htm

References


Authors Guild (2011b), “Found one! We reunite an author with an ‘orphaned work’”, body of article and comment by Stephen Bell, 12 September, available at: www.authorsguild.org/advocacy/articles/first-find-one-we-re-unite-an.html


About the author

Dennis Massie is Program Officer at OCLC Research and can be contacted at: massied@oclc.org

To purchase reprints of this article please e-mail: reprints@emeraldsight.com
Or visit our web site for further details: www.emeraldsight.com/reprints