Op Ed -- IMHBCO (In My Humble But Correct Opinion): Preservation, Yes -- But What Shall We Preserve?

Rick Anderson
University of Utah, rick.anderson@utah.edu
Op Ed — IMHBCO (In My Humble But Correct Opinion)

Preservation, Yes – But What Shall We Preserve?

by Rick Anderson (Associate Director for Scholarly Resources & Collections, Marriott Library, University of Utah; Phone: 801-721-1687) <rick.anderson@utah.edu>

Our work as librarians has always been the work of making difficult choices, but sometimes it seems like the choices we have to make are getting harder and harder. In this column, I’d like to talk about one that’s so tough we don’t even talk about it: how do we decide what information is not worth the trouble of preserving?

As people dedicated to collecting, safeguarding, and providing access to information, and as people with a social conscience generally, we’re loathe to say that any one kind of information is more worthwhile than another — we see value in classical music and pop music, in canonical literature and genre fiction, in perspectives from the mainstream and from the margins.

But as professionals, we also have to acknowledge the fact that we’re being paid to discriminate. We’ve always had to choose between resources that are “more relevant” and “less relevant” (given a limited budget, should I buy a history of Massachusetts or a history of Wisconsin?) and to some degree between “better” and “worse” (given that my library needs a history of Wisconsin and can afford only one, which one seems most reliable, thorough, and up-to-date?). But we’ve always made those decisions with the understanding that even if our library isn’t going to buy that history of Massachusetts, another library will. The book isn’t being lost, it’s just being cared for elsewhere.

But the question “What will my individual library collect?” is subtly but significantly different from the question “What must our profession preserve?” In a way, that question is actually easy to answer, because any answer will make us feel good: we must preserve this, and that, and the other thing, and no matter what the things are, there’s almost always a good reason to preserve them. But there’s another question that is just as important but much, much harder to face: what can we decide not to preserve? Let’s not be euphemistic here: this is a question that requires us to identify information that is, as the British put it, “surplus to requirements.” It requires us to identify books, journal articles, Websites, opinion pieces (yikes), recipes, oral histories, photographs, blog entries, musical compositions, and other documents that we are willing to let fade into oblivion, never to be seen or heard from again. Let’s be even more brutally realistic: this is not about deciding that it’s okay for my library’s copy to disappear — we’re talking about deciding what can be allowed to disappear completely from the human record.

Now, horrifying as that sounds, it doesn’t sound as bad as it could. Actively identifying information sources that can be let go at least requires the application of some measure of professional discrimination and training. It implies that we look at the whole array of what’s available (or at least a significant chunk of it) and make thoughtful choices about individual documents. Unfortunately, if we’re going to be realistic and hard-headed, we have to acknowledge that this is impossible.

Why? Consider this statistic: One fairly recent study found that the production of “new, stored information” increased at a rate of 30% per year between 1999 and 2002, and that the total amount of new information created in 1999 was five exabytes. That means that even if all the information professionals in the world united as one in a commitment to review and categorize all (or even most) the information produced in 2002, it could never happen. All of us probably recognize this, at some level of consciousness. But I’m not sure we all understand how monumentally impossible that task would be, and how microscopically tiny is the sliver of information output over which we have any influence as librarians.

At the risk of belaboring an obvious point, let me try to put these numbers into perspective: Five exabytes of new information were created in 2002. One exabyte of information equals one billion (that’s billion, not million) gigabytes. A home computer with a 100 gigabyte hard drive can hold the equivalent of 266,650 300-page books. Assuming a world population of 6.5 billion people, five exabytes of new information translates into 20,511 new 300-page books (unique titles) per person. In 2003, OCLC estimated that there were 690,000 librarians in the world. Of course, not everyone who takes care of information is a librarian, so let’s double that number. No, actually, let’s multiply it by ten, giving us a processing team of 6.9 million information professionals — this assumes that worldwide, one person per thousand is a member of the information profession. If we were to charge the information profession with reviewing, categorizing and caring for all of the new information created in 2002 alone, that would mean each professional would be assigned the equivalent of just over 19.3 million books. And that’s only for 2002. Assuming that the amount of newly created and stored information is still only increasing at a rate of 30% per year, for 2003 your assignment will increase to 25 million books, and the year after that it increases to 32.6 million.

In this scenario, each information professional would be charged with creating the equivalent of the Library of Congress — every year.

The obvious objection to the preceding paragraph is this: “Come on, Rick; you’re poking at a straw man. No one has ever said we can capture and take care of all the world’s information.” Granted. But how many of us realize how infinitesimal is the size of what we are able to capture and care for? Again: assuming — and this is an exceedingly generous assumption — that one person in a thousand is an information professional, that person can’t even come close to handling the rounding error on his share of the world’s information. Even if we allowed that only 1/100 of the information produced worldwide each year is worthy of an information professional’s attention, that amount of information is still completely impossible to handle.

And here’s why the straw man is relevant. In a previous column, I argued that we, as a profession, have a tendency to argue from value while ignoring opportunity cost — a tendency to say that we must continue doing X because X is valuable, while closing our eyes to the value of the things that don’t get done while we’re doing X.

What the ongoing, exponential explosion of newly-created information does is massively increase, in a mostly invisible but still urgently real way, the opportunity cost of everything that we do in the library. Every year, the cost of doing what we did last year increases at the rate of information growth, and that rate is already high and will only increase further.

So what does this mean for preservation? I think it means several things:

1. Painful as it may be to do so, we should explicitly acknowledge that the overwhelming...
The undertook a software project to help streamline the labor-intensive selection and ordering process. The result was the Integrated Tool for Selection and Ordering at Cornell University Library, or “ITSO CUL.”

Several years ago, staff at the Cornell University Library undertook a software development project to help streamline the labor-intensive selection and ordering process — and reduce the inefficiency inherent in learning and using multiple online systems. The goals of ITSO CUL were to assist library staff with the selection and ordering process, and bring together into one interface new publication records from materials vendors, the library’s profile and MARC records from the Library of Congress. In 2006, Cornell began working with a team from OCLC to further the development of ITSO CUL, and late that year, OCLC introduced WorldCat Selection.

“Paper Shuffling that Took Too Much Time”

The challenges inherent in a paper-based system for the selections and acquisitions were largely behind decisions that the McGill University library system and the Getty Research Institute (GRI), Research Library in Los Angeles made to implement WorldCat Selection.

McGill University, based in Montréal, has 13 libraries in its system, and Joseph Hafner, Associate Director, Collection Services, would be among the first to admit that a 13-library system can generate a lot of paper slips in selection and acquisition workflows. “We were using several vendors — and were still receiving paper slips from our European vendors,” explains Hafner. “These slips had to be sorted to determine the selector to which each belonged before they could go to our acquisitions team.

“The acquisitions staff then had to look online for corresponding records of the new items, and if none existed, they had to enter new records. It all amounted to a lot of rekeying and paper shuffling that took too much time.”

According to Ann Roll, Acquisitions Librarian, the Research Library at the GRI was balancing slips from its acquisitions vendors in both paper and electronic form. “This was difficult to facilitate,” notes Roll, “and slips often went unreviewed. We also saw an increase in slips from our European vendors — their allocations for books were being spent more quickly due to their favorable exchange rates.”

Meanwhile, interest in what Cornell was doing prompted Princeton University, in Princeton, New Jersey, to pursue WorldCat Selection, as it offered a way to work with Library of Congress resource file data more effectively.

I also realize that I haven’t exactly proposed a real solution to the problem of preservation in an environment of overwhelmingly explosive information growth. Ultimately, there may not be a solution. We may eventually have to let go of the whole idea of the library as a permanent repository, and flip the traditional collection model: instead of investing primarily in permanent collections, focus more on providing an effective portal to everything that’s available at a given moment. Not even the Library of Congress can handle everything that it really ought to. Why do we continue pretending that it — let alone the rest of us — can? 🥴