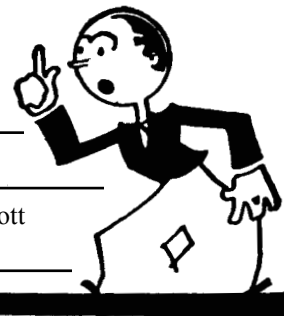


Op Ed — IMHBCO (In My Humble But Correct Opinion)

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 real-

istic: 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 study¹ 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 2002 — alone — was five exabytes. This 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 estimated² 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 re-

viewing, 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,³ 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 overwhelm-

continued on page 54

ingly vast majority of the world's documented intellectual output (what the **Berkeley** study called "new, stored information") is going to exist in the world only temporarily, and will eventually disappear permanently. This is no one's fault. It's simply the reality of a world where creating and distributing information has recently become easy and cheap while organizing and archiving information permanently remains difficult and expensive.

2. As librarians, we must set priorities ruthlessly. Knowing that we can't keep and care for everything that deserves to be kept and cared for, we have to reallocate staff time to the care of those documents that deserve it most and dispassionately take staff time away from objects and processes that deserve it even a little bit less.

3. Bearing in mind how tiny is the fraction of information over which we can actually exercise stewardship, we should rethink the principles we use to set those priorities. How can we tell whether a document contributes substantially to our institutional mission? What makes a document more worthy

of preservation than another one? Or, more to the point for each of us, what makes a document more worthy of *my staff's time* than another one? The documents that deserve it most may or may not be the ones we consider "best" — they are those that most effectively meet the needs of our patrons and help the library advance the priorities of the community it serves.

4. We must largely (though not completely) let go of our boutique model of both collecting and preserving. It's easy to leaf through a publisher's catalog and find titles that look interesting. It's easy to decide that the damaged book I see in front of me right now deserves to be repaired. It's hard even to comprehend, let alone honestly confront, the huge and growing opportunity cost imposed by directing time to those activities.

I realize that this whole column tends to conflate the issues of preservation and collection development. But that's partly because the connection between them is so intimate. Preservation is basically the enforcement arm of collection development — it's the mechanism by which we make our collecting decisions stick. Decisions about collection development are necessarily preservation decisions, and vice versa.

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? 🌿

Endnotes

1. **University of California, Berkeley School of Information Management and Systems**, "How Much Information? 2003." October 27, 2003. <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm> (accessed October 16, 2008).
2. **OCLC**, "Libraries: How They Stack Up." 2003. <http://www5.oclc.org/downloads/community/librariesstackup.pdf> (accessed October 16, 2008).
3. **Anderson, Rick**. "On Knowing the Value of Everything and the Cost of Nothing." *Against the Grain* 18, no. 6 (2007): 44-45.