

My Three Wishes for Digital Repositories

By Terence K. Huwe

My interest in digital repositories goes back to the dim and venerable year of 1997. At that time, research libraries were already pretty far along in exploring standardized general markup language (SGML) and Web-based solutions to information management. Flat-file UNIX databases running on CGI had emerged as a terrific way to build resources, because they presented a comparatively low hurdle to getting things up and running. There was a growing awareness that our user communities really wanted solid scholarship on the Web and that it was possible to gain their "attention" by quickly rolling out new repositories. The more proactive among us started homesteading a new role for the digital library with "born-digital" resources. Basically, it was records management on steroids in business firms; at universities, it was a whole new "hook" for involving the faculty, not to mention raising funds for new library initiatives.

Fast-forward a few years, and we see that we've come a long way. The Open Archive Initiative led to applications like the Online Archive of California (OAC), which offers access to high-quality collections of text, image, and sound files. The OAC is not the only show going, but it's a popular one, as witnessed by New Mexico's adoption of its architecture. Another local initiative, the California Digital Library's eScholarship program, took flight and grew quickly. As an early adopter, the library I manage jumped in and pitched the repository to the faculty at Berkeley, with great results. We now oversee four working paper series that grow in spurts and starts, and we've digitized a full run of our flagship series. The staff members have also offered "information counsel" to other journals published by this institute. It's a great thing to practice what you preach—all the while playing the role of expert consultant in areas where smart people can safely be assumed to be clueless. Plus, it worked: One of our prestigious law journals, *California Public Employee Relations*, adopted e-commerce to sell its print editions. It also sells a digital version via the Berkeley Electronic Press (<http://www.bepress.com>).

Everything's going just great, right? Not so fast: Yes, the profession has gained ground in digital repository management, and those skills are becoming more

widespread. But I see an overall environment where digital librarians excel in the predictable ways like metadata creation, quality standards, and finding aids—certainly achievements we can be proud of. Yet at the same time, I also see a shortfall in the "vision" department. I can find three areas within the sphere of digital repositories that need work. The first two pertain to information architecture, while the last one pertains to taking action. And action is important, because digital repositories will play a central role in research environments. How we integrate (or fail to integrate) them into our information ecologies will be instructive as a test of how much we can grow—not only in collection development skills, but also in access strategies. So here are my three wishes.

Bridge the Archipelagoes

Many of us are lucky enough to work for libraries that maintain top-notch Web sites. A good library Web site is a gateway to any sphere of knowledge. The University of California-Berkeley's Library Web (<http://lib.berkeley.edu>) is as good an example as any, and it has a lot of company at peer institutions. However, there's one thing we do not see enough of, as far as I'm concerned. We need to see far more integration of born-digital, or "built" content with traditional library collections and services. As Marshall Breeding recently pointed out (see *CIL*, January 2005, p. 28), integrated library systems could do more of this, but they are playing catch-up in the design process. Until the ILS or other utilities can offer a better approach for linking all e-resources together, our best tool is the library Web site itself.

Because we care so much about the challenge of good usability, we want to believe that we are doing enough, but we are not. We need to go further, because our patrons expect—to a much greater degree than just a few years ago—a rich, intellectual context on the Web. Despite all of the advances in digital repositories and full-blown archives like the OAC, the strongest threads of integration are the links someone hopefully remembered to put on a Web page. The links should be anywhere from one to 10 clicks away from the question at hand.

Wait a minute: We link e-journals within the OPAC, right? Sure, and that's good and getting better all the time. But it depends on the user remembering to go to the OPAC in the first place. What about Web-hosted databases of e-journal titles (with metadata included), working papers, or e-books? These can be made searchable by access and coverage periods, and they can be listed right there for the researcher. Not bad—Berkeley's Library Web has this feature. Put the two together and you may be OK. Just make sure you keep track of moving walls in

e-journals, growing repositories, industrial-strength scanning of collections, and other fast-moving ventures that mark our daily lives. Web databases of e-journal titles coupled with OPACs are important, but for them to be effective they must be usercentric. Indeed, whenever I compare the research processes people employ with the Web tools they must use, I am reminded that the best Web sites are crafted pieces of work and bear the mark of their makers.

Library Web sites are crucial gateways to content beyond the OPAC, and how we manage their usability to link diverse resources has become one of the most important design challenges we face. I have two suggestions for anyone who is responsible for Web links to families of documents, books, finding aids, and other media. First, use your top-level browser real estate to show links to all of the born-digital content you manage in addition to the library catalog. Second (since you won't be able to fit it all on the top level), use the phrase "More ..." as linked text to deeper levels of finding tools.

Enhance the Context

Repositories create new opportunities for us to demonstrate how digital libraries provide well-crafted intellectual context, which can energize effective research habits. But how can we extend that context even more effectively to our repositories, which are meant (foremost) to be "persistent" locations for the long term? Based on my own experience, the answer lies in page design.

Academic digital repositories need a new visual strategy that shows readers other relevant links and documents, at first glance. My second wish, then, is for dynamically generated pages that show related documents—much as we've grown used to in using commercial Web services like Netflix (<http://www.netflix.com>). Pages that point at related resources would add greater context to each document with its specific contents. The best place for this feature would be on the abstract, or entry page of a document or digital object. Here's an example where this approach would help.

In California, project labor agreements (PLAs) are a very big deal. Without getting too technical, PLAs are contractual agreements where all employees working on a major construction project (a power plant, college campus, or hospital) enjoy similar pay scales, benefits, and employee rights. This streamlines employment relations and enables the real work to get done faster. As a topic, though, PLAs are political hot potatoes on both sides of the aisle, and they generate massive debate and controversy. Whenever my institute publishes on

this topic, we see an avalanche of downloads. It would be ideal to have a dynamically generated list of bibliographies, Web links, and commentary appear with the new paper.

The skills required to generate this kind of feature exist right now — and I hope the idea inspires some swift code writer to give it a whirl.

Everyone: Get Involved

Is anyone out there managing or contributing to a digital repository? It seems like one of us could be. This is a fair concern, based on feedback I get when I make presentations about repositories at Computers in Libraries, Internet Librarian, the Special Libraries Association, and within the University of California Libraries (which employ nearly 500 librarians). If I ask for a show of hands when I ask this question, I rarely see any go up.

Therefore, my third wish is for more of us to get serious about creating and using repositories as key elements in our overall information ecologies. This idea shouldn't be "rad": After all, special librarians have recognized the strategic value of "gray literature" for decades, generating forms files, brief banks, and other forms of knowledge management. Digital repositories are the killer app for gray literature. In our case, at the University of California, the eScholarship repository reduces the labor to a matter of "click to add" and allows the professional to think about metadata, access, and outreach issues. It's really simple to use, and librarians are the right folks to create and manage collections in this arena. I wish more were.

Google's stated desire is to be the "brain" of society, available anywhere and at any time, which leads to many insights. My favorite is simply this: No matter how any knowledge resource is born, it can be digitized, and it gets transformed in the process. The real potential of digital repositories lies in their ability to capture ephemeral documents and combine them with more familiar formats. When it comes to built content — the stuff our organizations generate — repositories can fill gaps, expanding the gray matter of strategic information. But like brains, they need to be put to use.

My three wishes for digital repositories can be boiled down to one wish for the profession: that we take the lead in each new digital platform as it appears and that we bring our user-centric approach to bear on each new tool. Digital repositories and full-scale archives are strategically valuable tools that aren't focused on books or journals per se, yet they add a lot of long-term value. And

our long-term future, like it or not, is linked to this platform and to its successors, which are surely "coming soon."