

Scanning Is Still Hip High-Tech

By Terence K. Huwe

As a profession, we have navigated some pretty rugged terrain, particularly if you like to trace the on-line "dawn of time" to the late 1970S like I do. When I enrolled in library school in 1979, 300 baud-rate connections were cutting-edge. Since then, and in stark contrast to many friends of mine who work in professions like law or accounting, I find that I learn how to use at least two new technologies per year. Of course, this is standard procedure in our profession, and our long-term experience provides plenty of fodder for that perennial favorite at the Computers in Libraries conference, Looking at Dead & Emerging Technologies (http://www.infotoday.com/cil2004/wednesday_evening.shtml). So it's all the more fun to take a closer look at what's new in the ultra-hip, ultra-cutting-edge world of scanning.

Scanning is far from dead. It remains a fundamental skill for digital librarians. Our collections, even when you subtract the percentage of copyrighted material, still have vast amounts of print treasures that are waiting to go digital. The richest of us can contract whole collections to jobbers who will, for a tidy fee, create searchable PDF documents that fit very nicely into digital libraries as objects within a persistent link. Images? Hey, the only real difference between an image file and a text file is the need to guarantee a "high rez" archival copy. Both files need metadata, which can be added to the digital object in fancy repositories, or simply typed as "meta" links in garden-variety HTML pages in a no-frills Web environment. Well-funded projects can be turned around in a matter of weeks, delivered on a set of CDs, and uploaded forthwith. That's the fancy way to go.

But not everyone has the wherewithal to farm out scanning work. Let's face it—you don't always get every grant you apply for, but that's not a reason to delay creating digital resources by scanning. We have grown accustomed to hearing about our largest and most respected research libraries' latest digitization projects. Some, like the Library of Congress' American Memory project, are truly monumental in scope (<http://memory.loc.gov>). But the glamour of the monumental can obscure the other end of the spectrum, where you, as a tech-savvy collection specialist, can create a modest-sized, manageable digitization project. This kind of cottage industry can do more to enhance your reputation as

an innovator than many other activities. And the new generation of scanners makes it easier than ever to get started.

Scanning's Strategic Value

Nowadays, more and more books, journals, and other library materials are "born digital," and our jobs are increasingly becoming devoted to crafting access strategies. But what about the still-relevant corpus of knowledge that was not born digital? The Northeast Document Conservation Center continues to offer a School for Scanning conference (<http://www.nedcc.org/prelfs/prgen.htm>) to help professionals learn best practices for conversion projects, from start to finish. This hands-on approach is immensely empowering, because it can help you master the details of OCR, metadata creation, and archival-quality digital objects. Indeed, the skills associated with print conversion will remain in demand for some time to come and are the foundation of digital library development know-how.

But a new generation of scanners brings more power to your hands without so much formal training. These products are reasonably priced and are much more reliable than their ancestors, and they enable you to get started with digitization without depending on grantsmanship. Here are three mission-critical reasons for re-evaluating your digitization plans, in light of the new technology.

1. **Portability.** As is the case with so many evolving hardware products, scanners have gotten smaller, quicker, and lighter. This can be handy if you are conducting research in the field or are just building a database of articles for personal use. Sometimes it's preferable to set up a scanning session in an out-of-the-way place so that you can examine collections of documents, photos, or ephemera in their "native" habitat.
2. **Scalability.** Anyone who lived through retrospective conversions of card catalogs knows what industrial-strength automation is like: huge, incremental, and slow. However, a lot of really interesting digital libraries lie waiting at the other end of the project spectrum: not too big, but high in value to researchers. For example, my library has the most complete collection covering the "California School" of industrial relations— the cooperative history of labor and management that grew during World War II, particularly at the Kaiser shipyards that built "Liberty ships." Scanning this collection and creating a digital library could be achieved in months, not years.

3. Marketing. In an era of instantaneous "googling" and ubiquitous Web information, we need to distinguish our collections and services by emphasizing their unique qualities. Many of these unique resources were not born digital. Digitizing them creates an instantaneous marketing opportunity to distinguish ourselves, and our library's imprimatur, from the "semi-decent, sorta OK" world of the open Web.

Meet the New Scanners

Hewlett Packard's 4670 is a light-weight, flat panel scanner that you can carry with you into the stacks or even to remote locations. It's simple to use, and, in contrast to the bad old days, the OCR software is superlative, producing clear text copies in one scan. It will drop a finished scan directly into Microsoft Word, from which you can create a portable document. Add a metadata header, and you've got a digital object in your hands. Locally, we've used this scanner to create digital images of posters, newspapers, and ephemera with unique historical value.

Hewlett Packard's unit is a "best in class" entry that most librarians will like, but there are smaller units that aren't even designed with the desktop in mind. The DocuPen, which was evaluated in *The New York Times* Circuits section (July 8, 2004, p. E8), is so compact that you can carry it with you wherever you go. Moreover, it's designed to help organize and keep track of scans, along the lines of those less-facile digital rolodexes. Its weakest point is the resolution quality, which comes in at 100 to 200 dpi (dots per inch). This pretty much bars creation of archival-quality scans, for which 300 dpi is considered the lowest acceptable number (with 600 dpi preferred).

For the large-scale scanners among us, there's Fujitsu's ScanSnap FI-5110-EOX, which is a moderately priced but high-powered machine aimed at the records management market. This scanner will copy 50 pages at a time and convert them into *searchable* Adobe PDFs, ready for uploading. Fujitsu envisions this kind of "almost industrial strength" scanning as a solution for paper-intensive industries and services, like law and medicine, where meticulous records and multiple copies are required.

What About Scanning for Smaller Libraries?

In each of these cases, the scanners aren't being actively marketed to librarians. There are probably several reasons, not least of which is the fact that large

funding agencies are still paying for library digitization projects. The libraries that can handle large-scale digitization also tend to have sizeable staff complements and can integrate new work flows more easily than smaller libraries. But those small libraries—perhaps yours—may have special collections that could be readily scanned with the new generation of technology. Ironically, smaller libraries are often hit the hardest by budget cuts, as anyone in California can attest. But tight budgets shouldn't create roadblocks for creative work. Indeed, in times when the government faces shrinking budgets and firms contemplate outsourcing, creative thinking is the gateway to survival. Here are three strategies I can think of to leverage your current collection, with the new scanners that are cheap and easier to use.

1. Survey the Collection. What stashes of photographs, memorabilia, or other historical materials do you have? What printed records do you have that will have historic value? Recent historical research about the antebellum South relies heavily on court transcripts to give the most accurate description of slave life before the Civil War. Is there something of similar value lying dormant in your stacks? As information professionals, it is our job to bring these materials to light—and to keep them accessible beyond today's fads or scholarly fashions.

2. Use Newly Digitized Collections to Be Distinctive. Library Web sites and blogs, at their best, offer visitors distinct experiences and unique resources. However, all too often we stop at the mere description of our print collections, which doesn't tell the visitor all that much. Select a small, manageable portion of your unique materials, scan them, and upload them: This will make your collection more eye-catching.

3. Start Small, Grow with Time. Success begets success, nowhere more so than in grant writing. Oftentimes a tangible resource, like digital collections in nascent form, can help win really big bucks, especially for persistent grant writers who establish a track record. Creating a prototype digital library from scratch may yield a new revenue stream later on, or it may help you build a relationship with a consortium, based on new awareness of what you have to offer.

The Future of Scanning

As always, the human side of digital library development involves some fairly nitty-gritty work—like scanning. But cumulative effort sustained in small bits and pieces over the long term can yield big results, especially if you match small-scale digitization with a well planned strategy for persistent access. The new

generation of scanners, with their ease of use, portability, and power, present us with a new opportunity to expand our role in the digital sphere.