Best Practices - Scanning

In early November 2007, Amy Zimmer, Library Technician II at the Colorado State Publications Library posted to the Colorado Libnet listserv the following inquiry regarding the use of scanning technology in libraries:

The State Publications Library (part of the State Library) is currently in the planning process for digitizing our collection. We are a very small library with a staff of four, and we have about 20,000 print titles needing to be scanned. For libraries who have done this, I have a few questions:

1) We are considering outsourcing the actual scanning. Can anyone recommend a good company to work with?

2) What are the advantages/disadvantages of scanning in-house vs. outsourcing?

3) What equipment can you recommend for in-house scanning?

Any help you can provide is most appreciated.

Thank you!

-- Responses--

I had experience with a Minolta PS7000 overhead scanner. Reference with specifications/features:
http://kmbs.konicaminolta.us/content/products/models/ps7000.html

I've also attached a photo from my previous library showing this scanner unit, so you get a better sense of size/scale.

For our Service - Pros:

- Could accommodate a variety of publication sizes.

- Was easy to use (minimal training required); hourly students ran the scanner.

- Fit easily into a production workflow that involved electronic delivery/distribution of nearly 30,000 articles/book chapters (more than 185,000 pages) every year. Scanner was in operation 6-8 hours each day.

- Corrected for curvature of pages (important for our situation, which often involved bound journals).

- Equipment was leased, which meant we didn't have the full hardware investment.

Lease arrangement included servicing. Leasing also meant that we could upgrade to new equipment when the time came.

General - Cons:

- Physically large unit, with an attached computer, took up a full table.

- ONLY scans in B&W. Promised upgrades to 8-bit grayscale as yet NOT provided (company rep could confirm). Color imaging NOT in the plans for this line of equipment.

- No OCR or conversion of image to text was built into our solution.
This was purely capture to PDF and shipping it off (digitally). This meant we had no ability to search on text within files, and no indexing of content. This text conversion or indexing activity WAS possible from the images captured during scanning, but would have required third-party software and additional processing/time/quality assurance and control measures. For our service, this functionality just wasn't needed. Our service was all about article delivery. Your situation is likely different...

Now, about your question on in-house vs. outsourcing of this kind of work...

In-house - we did this work in the library because it was a service operation processing thousands of pages annually, and would continue to do so for the foreseeable future. Over the long haul, with student employment, the work was less expensive than outsourcing.

However, WE had to develop the entire workflow and adjust the workflow when changes in hardware/software/systems would occur. This was a stressful endeavor, because the demand for article delivery never abated. In-house IT support was crucial to the success of the service.

Outsourcing - had this been a sizeable, one-shot retrospective digitization project, we would have seriously considered outsourcing. No need to develop workflow; just send 'em your content and *snap!* -- receive data DVDs filled with images, PDFs and indexed content for loading into your asset management system! Outsourcing is generally faster.

However, outsourcing typically costs a lot up front.

The way I look at it--either way, you're paying *someone* to do the work.

On the question of general digitization/scanning/production work...

I would have to know more about your framework, your restrictions, your end goals for your digitization project to recommend any flatbed scanners. I do have experience

with these as well (including document feeders). I also have experience with 35mm slide scanners with hardware-driven image enhancement functions. I assume your content isn't film, though...

Jim Duncan
Colorado State Library

I hope you would be bombarded with the response: CDP! which is now CDP@BCR!

http://bcr.org/cdp/index.html

We are so fortunate to have a premier organization that pioneered best practices in digitizing--and it is right in our own region. These best practices are key to doing these projects correctly which I strongly urge you to follow. Best practices are key to many parts of a good digitization program, not least of which is that the material will be digitized in such a way that it will still be accessible in 10-20 years.

This organization has gone through some transformations, now resting with BCR, but it should be a good source for the information you need.

Carl Stewart
University of Colorado, Boulder

I saw your posting to Libnet about scanning and the pros and cons, and thought that I would get in touch. Recently, BCR merged with the Collaborative Digitization Program (CDP) and we offer training and consulting on the subject of scanning and digital project management.

Leigh Grinstead (lgrinste@bcr.org) is the Project Coordinator for the CDP@BCR and she would be the perfect person to give you the answers to these questions. I would recommend giving her a call when you get the opportunity, and I am sure that she can help you out.

Regan Harper
Bibliographical Center for Research

I have found the following books to be very useful:

Kenney, "Moving Theory into Practice" (2000)

Stielow, "Building digital archives, descriptions, and displays : a how-to-do-it manual

for archivists and librarians" (2003)

Also, are you aware of the Colorado Digitization Program?

http://www.cdpheritage.org/

Anectdotally, if you can afford to spend about a dollar per page, outsource it. But the above resources are good at outlining the questions you should be asking to make sure you arrive at the best decision for your repository.

Jordan Steele
Biddle Law Library

You might start with the Wikipedia "Digital libraries" page:

http://en.wikipedia.org/wiki/Digital_library

Either in-house or outsourcing you have to work through project design detail by detail, probably starting with an extremely precise and technical description of your goal, then working back through the most efficient means to reach it. The obvious advantage of in-house is that you have more control at every step of the project. The obvious advantage of outsourcing is that you do not need to invest in equipment, people, training, and supervision at every step along the way.

On the assumption that other Colorado libraries also hold state documents, this sounds to me like a project on which it would be natural and cost-effective to organize a multi-institutional consortium -- in which case a third choice might be to centralize scanning and some other functions in a consortium-supervised facility.

Sidney Huttner
University of Iowa Libraries

The Collaborative Digitization Program created a great deal of info about digitizing that's available online; BCR is currently overseeing the program, complete with training workshops, so I suggest you contact them.

Becky Lintz
Colorado Historical Society Library

I was going to reply earlier, but we're in the same boat as you are. Our state documents librarian retired back in May, and we're having a dickens of a time trying

to get that position filled. Until that is done, we're holding back on planning anything for scanning any of our docs. We are the sole depository for state docs in Wyoming, and have about 104 units of 7 shelves each. The University of Wyoming and the Wyoming State Archives selectively keep some items.

The Wyoming Historical Society worked with Northern Micrographics to digitize the Annals for the Western Trails Project. We're using PTFS in Bethesda, MD to digitize our newspaper microfilm. We do a lot of scanning in house, but right now it is just me and a $6000 Canon scanner, and I don't have time to digitize all that we want to. The Wyoming Water Development Commission checks out some state and federal reports and gets them scanned at a local copy shop. The scanning is ok, but sometimes the pages are out of order on the PDF files, they are image only (I use ABBYY FineReader to OCR them), and sometimes I have to crop the pages.

We bought a Canon DR-7080C several years ago to scan community planning documents in partnership with the Wyoming Business Council. It has an 11x17 flat bed, and does simplex and duplex scanning up to 11x17 paper, in black & white, grayscale, and 24 bit color.

Montana State Library http://msl.state.mt.us/news/dltp.asp recently created a bit of a stink about their plans to close to walk ins and digitize everything.
They plan on scanning in-house.

You could also post your questions to GovDoc-L or DigLib or ImageLib.

I assume you've checked out the project management section of the CDP@BCR (or whatever they are calling themselves these days) at
http://cdpheritage.org/digital/index.html

Erin Kinney
Wyoming State Library

I have suggestions for two people that you can contact who have been involved in digitization projects:

Shaun Boyd:
Douglas County
History Research Center
Philip S. Miller Library
100 South Wilcox
Castle Rock, CO 80104-1911
(303) 688-7730
Fax: (303) 688-7715
e-mail: sboyd@dclibraries.org

Allaina Howard at the National Snow and Ice Data Center Allaina Howard
allaina.howard@nsidc.colorado.edu

Alice Kober
Arapahoe Library District

My experience comes almost 100% from working with the CDP, online at
http://www.cdpheritage.org/cdp/index.html

They have a lot of helpful information.

Clara Burns
Front Range Community College Library