Introduction

For the past three decades, the United States Departments of Health and Human Services and Agriculture have presented *Dietary Guidelines for Americans*. This document "provides information and advice for choosing a nutritious diet, maintaining a healthy weight, achieving adequate exercise, and 'keeping foods safe' to avoid food borne illness" (US Department of Agriculture and Department of Health and Human Services, 2005, p. i). Key to the guidelines' success is mandating effective communication through nutrition facts and food labels. These tools "provide information that is useful for implementing the key recommendations in the *Dietary Guidelines* and should be integrated into educational and communication messages" (US Department of Agriculture and Department of Health and Human Services, 2005, p.3). Yet, despite much effort, the complexities of food and nutrition communications remain confusing to many American consumers (Grimes, Riddell & Nowson, 2009). So it is with selecting complex hardware for the modern digital library.

This paper describes a decision-making structure for investigating, selecting and installing digital library hardware at a large U.S. academic library. It shares findings related to the challenge of exacting important and arcane engineering detail from vendors and the necessity of parsing the content of their advertising and technical copy. In our descriptive case study, we point out the key components of a decision-making schematic for the selection of scanning hardware.

Literature Review

Scanning or digital imaging has been part of the library community for nearly two decades. Writing in 1996, Kenney and Chapman offered a theoretical overview of the key concepts, vocabulary, and challenges associated with digital conversion of paper and film based materials. Of particular concern is their overview of the hardware and software associated with implementing a technical infrastructure to support a full imaging program (Kenney, Chapman, 1996). Lee's *Digital Imaging: A Practical Handbook* follows in 2001 and again provides some guidance on the selection of hardware. His advice on setting up a series of questions on "ease of use" and determining "effective resolution and bit-depth" are valuable and now time-tested (Lee, 2001, p.51). He also offers a nice survey of scanner options noting, "one of the major limitations of the digital cameras…is their inability to handle large-format documents" (Lee, 2001, p.58).

Subsequent to these early writings with their focus on the complete digitization workflow process and high-level overviews of associated technology, came a spate of technical papers on imaging standards. Writing in 2002, Burns et al highlight the necessity of understanding the International Organization for Standardization (ISO) standards for measurement of image resolution. They contend that "identifying conditions that lead to measurement errors" and reducing the estimation factor mitigates against bias and variation (Burns, et al, 2002, p.4). In a follow on paper, Burns and Williams focus on using the slanted edge analysis to improve evaluation of image resolution for scanners (Burns & Williams, 2002). Often referred to as noise, unwanted image fluctuations has been the focus of several papers (Zeise, et al, 2007 and Baird, 2007). Then in 2006, Stelmach warns the community not to be seduced by manufacturers and their calm assurances of image quality. He suggests that cultural institutions should employ statistical process control (SPC) strategies “to monitor the output quality of their imaging workflows” (Stelmach, 2006, p.1).

Williams predicts the main thrust of our work quite plainly when he states,

> “And so it is with the claims of the capture performance of digital imaging devices. One is beckoned by a cacophony of vendor specifications. Loud and confusing (and unregulated), they are, ironically, seductive” (Williams, 2003).

While Williams and others offer much needed specification, they also give a false impression of the amount of information available to libraries for making informed purchasing decisions. One can find a decade’s old report (Boss, 2001) that provides guidelines yet, given the massive advancement in technology, it suffers merely from its publication date. Outside the scholarly literature, one is able to locate some general informational brochures and tips and tricks. An example, readily available on the WEB, is Jenkins’ *How to evaluate a scanner* (Jenkins, 2008). In sum, while there is a wonderful array of technical papers available, there is a paucity of contemporary, targeted, general audience-based papers for the selection of large format documents scanners. We hope to remedy this by providing a decision-making structure for evaluating, selecting and installing digital library hardware.

Case Study

The Michigan State University (MSU) Libraries plunged headlong into the large format, high-end scanning arena in the spring of 2009 when staff accepted the challenge of digitizing one hundred years of unique newspaper broadsides. Grant funding for the digitization of the University's "Janet A. Ginsburg Chicago Tribune Collection" came to MSU's School of Journalism from the McCormick Foundation in 2008 and mandated the physical restoration, storage and digitization of the collection.

The collection is a five thousand-page selection of stunning daily Chicago Tribune newspaper graphics dating back to 1898. Many of the pages feature beautiful four-color images and all are a product of the groundbreaking Tribune technical and design staffs. The emphasis of the collection is the newspaper as graphic art.

The job of digitizing the newspapers fell to the Digital and Multimedia Center (DMC), the MSU Libraries' in-house digitization operation. The task was complicated by the fact that the decision had been made to digitize the pages, all of them at least A1 (33" x 24") in size, at the high-archival standard of 600ppi (pixels per inch), 24 bit color.

At the time, many in the imaging world considered a 600ppi resolution to be unnecessary for large document scanning. It was thought that this over sampling was simply a waste of space. The Library of Congress (LOC) best practice guidelines suggested a 300dpi minimum and warned that doubling "may not produce a visible image quality improvement when the spatial and tonal resolutions are already sufficient" (Library of Congress, 2010). We could not, however, rely on LOC guidelines or on Cornell's well-respected Quality Index Formula, (Cornell University) as the main thrust of our project was not text or photographs, the subject of most of the best practice guidelines. So, in consultation with the MSU Archivist, the decision to scan the MSU Tribune Collection at 600ppi held. It was also decided that shooting in 600ppi, 24-bit color would be the best way to truly capture the incredible Tribune color graphic images. Any future digital derivative (jpg, pdf, etc.), it was believed, would certainly benefit from being based upon a high resolution, color, raw tiff file.

The DMC is a relatively new imaging shop and when we received word of the technical specs we would be expected to follow as part of the Tribune project, we had many concerns. First, our base equipment for large format work was a very good planetary scanner, capable of a 400ppi resolution scan at A1. When asked, we call ourselves a "boutique scanning" operation and have the capacity to scan upwards of 250,000 pages per year of MSU Library collections, using a combination of planetary, sheet feed and flatbed scanners. We worried about venturing into a realm where 600ppi was to become our norm, especially when applied to large document formats. Fortunately, we soon learned that the grant funding for the Tribune project included money for the purchase of hardware and software to support the digitization of the newspapers. It also soon became apparent that a great deal of money would be necessary to take us into a rarefied atmosphere where imagers accurately scanned A1 sized documents at a true optical 600ppi.

Choosing a Scanner

Our search for the perfect scanner began as most library quests begin, with the formation of a committee. Staff from DMC and the Library's Systems unit was brought together to establish a set of criteria for selection, determine a research process and pick hardware and software by consensus. Obviously, the most important part of the committee's charge was to select a scanner capable of meeting the needs of the Tribune project. Since any equipment purchased would become the property of the MSU Libraries', projected future applications for DMC, the MSU Libraries in general and the MSU campus as a whole were also discussed. We knew that we had a unique opportunity to become a very high-end scanning shop on campus and we wanted to make sure that whatever hardware and software we bought would be able to serve a wider, ongoing audience. This meant that the device must be flexible. It must, of course, be capable of scanning large documents at high resolution, but perhaps also very large documents such as maps and at the same time, our usual library trade of books and journals.

Production in general also must not be ignored. Any device under consideration would need to offer production speeds, both in scanning time and throughput, which were reasonably fast based upon document size and desired resolution and be consistent with our high volume/high quality workflow. Since our workforce is primarily made up of undergraduates, ease of training was also an important issue in the face of production demands. The device and its software must be relatively easy to use and a diverse crew of eighteen to twenty-two year-olds must be able to acquire the necessary operating skills in a very short time.

In establishing our business case we were given a top end budget for scanning hardware and software. We soon learned that this was a low ceiling in the world we were moving into. In addition to the upfront purchase price, we were also authorized to contract for continuing technical support and software upgrades. Any contract we entered, however, must be affordable, as continued funding would eventually fall to the Library.

More important from a business perspective was to properly vette the prospective vendors and manufacturers. We wanted ultimately to work with a company who had been in business for a respectable length of time and had a proven track record in customer service and quality manufacturing.

We knew from the very beginning that our biggest project challenge was going to be the required 600ppi capture of the newspapers. As the diverse committee gathered, a simple yet strategic set of questions was shared with all members with the goal of driving the committee to consensus on the priority elements (see Table 1).

Table 1: Strategic Question Set

What is the minimum information needed to make an informed decision?	Eg. Capture resolution, document type accommodation, speed, output image quality, cost, physical footprint
What (more) information would we love to have to make a decision?	Eg. Software, interoperability, customer support, accessories, usability
What is the cost of collecting answers to these questions?	Eg. Travel, staff time
Can we assign a priority level to the arrived at elements?	Employed a proportional voting procedure

It is important to note that in the early stages of our discussions we were afforded the opportunity to candidly argue for the 400ppi resolution standard accepted by most imagers. Even a cursory product review revealed that machines capable of scanning A1 at 600ppi were scarce and we knew that the immense file sizes (projected to be 600 to 750 mgb per file) created would be difficult and maybe impossible to handle given our current Library systems infrastructure. Fortunately, the library's IT staff successfully overcame these infrastructure limitations and wiser heads prevailed that kept our charge of capturing at 600ppi in place.

It became evident in our deliberations that our goal must be the capture of a "true optical" image of what only the camera sees without excessive image treatment or "interpolation". Interpolation relies on the computer to "see" and decide what should be added to complete an image. Excessive image treatment or correction can add undesirable noise to the process. This became our most critical benchmark. Our aim was to avoid devices aspiring to 600ppi by up-scaling or adding (interpolating) information which the software believed to be a correct rendering of what the camera it supported could not really capture. The mission of the committee was to shoot an extremely high quality, 24 bit color 600ppi resolution, true optical digital picture of an eighty-year-old A1 size newspaper page with four-color graphic images, on a single camera pass. Preservation would mean an attempt to create the most accurate, faithful true copy possible and to carry that image forward through time and space.

After completing our preliminary discussions and with our specifications and guidelines in hand, we began our investigation of available products, distributors and manufacturers. Unlike most libraries, MSU is fortunate to have an engineer on staff, who brought to the table a true understanding of the science behind the device we sought to buy. As our engineer carried out his product search and made calls around the country, he found that most large format scanners were unable to meet all of our high priority criteria, including image quality, capture resolution, document size, preservation handling, production speed, and ease of use—all of the vital project points we had established through discussions based upon our Strategic Questions discussion grid (see Table 1). As previously stated, the 600ppi image capture variable became a driving force. The initial investigation revealed that the only way most scanners could achieve a 600ppi scan of an A1 image size was by digitally stitching two images together or by using software to interpolate image data. In the end, we found only two manufacturers who claimed to be able to deliver a true optical, 600ppi digital image of an A1 document, without interpolation or significant image enhancement, at reasonable production speeds. We shall call them Firm A and Firm B and offer our experience as a guide for libraries and librarians in the selection of very complex and very expensive imaging hardware and software.

Key Considerations

Both firms were foreign manufacturers with American distributors and sold planetary scanners with camera heads containing linear array CCDs (charge couple device). In Firm A's machine, the single 14,000 pixel CCD camera traveled across the full fifty-inch width of the scanning bed with a maximum page height of thirty-three inches (A0 size). In Firm B's device, a camera with two 7,000 pixel CCDs moved top to bottom in a sort of butterfly wing pattern, for a maximum scan area of thirty–three inches (width) by twenty-four inches (height). Both claimed to scan A1 size documents at 600ppi and Firm A's machine would also scan A0 documents at 400ppi. Both included image correction and enhancement software for post-scan image treatment in their accompanying software packages and both devices were about the same price.

This late stage of our search included interviewing existing customers of both companies, discussions with outside, impartial imaging experts, ongoing talks with representatives from both companies and securing product information from both in the form of shared software, Webex and online demos and the sharing of scanned test targets.

One of the best things to come out of our scanner search was the kindness and support we received from our colleagues. We contacted folks from around the country who had purchased the devices in question and found them more than willing to discuss their experiences with using the scanners, training staff and the responsiveness of the distributors and manufacturers, after the purchase. Their help was invaluable. Due to internal budget restrictions our traveling was limited. We did make several trips within the state of Michigan, but the rest of our peer investigation was done over the telephone and through e-mail. The competing vendors were also limited by budget in the amount of traveling they were able to support. Expecting them to visit MSU with their huge scanners in tow seemed to be an unreasonable expectation. We decided to rely on telephone, Webex and online demos to learn about their products and both companies were very good about supplying us with information. We were able to try all of their software applications online and to get a pretty good idea of how the devices worked, without having seven-foot-long scanning beds shipped to East Lansing.

A most interesting part of this vetting process, and something we had not anticipated, was the sharing of "test scans". These generic test sheets are gradated line and color charts used for precise scanner calibration and evaluation. We sent the same physical test target to each vendor and received back CDs with the resulting test scan showing how each was able to render the card. We also conducted on-site test scans during our visits to two facilities within Michigan so that we could directly control and equalize all variables. The main thing to be learned from these types of examinations is how well a device separates color in images and the amount of noise introduced at high resolution digitization. A careful examination of test images (known as visual literacy) can show color shifts and bleed and most importantly, how well a device resolves the smallest picture elements where the pixels start to blur or break up (Burns, 2007). We also sent the companies examples of our scans of journals and books to find out the effectiveness of their post scan image treatment software. Having test scans and other imaging samples from both manufacturers for side-by-side comparison proved to be incredibly important in the final stages of our selection process.

Another interesting discovery was the lack of exacting and accepted measurement standards in the digital imaging industry. As our engineer continued his research, talking to vendors and their customers, he found that there was little consensus and no accepted benchmarks for testing, marketing and presenting specifications. This has a great impact upon how any given sample image might look and makes the comparison of devices incredibly difficult. This is especially true in discussing how an image is to look at a specific optical resolution, and at various scanning speeds and how much image noise was really being generated. Some image noise is generated by the CCD itself, which increases at higher scanning speeds and with higher gain or exposure levels. (Zeise, et al, 2007).

An acknowledged expert in digital imaging, Don Williams of Image Science Associates (www.imagescienceassociates.com), kindly gave us a short course over the telephone on the arguments surrounding imaging evaluation and candidly explained his take on the strengths and weaknesses of the devices we were considering. Williams is a strong proponent for the creation of measurement standards and practices for the digital imaging industry.

By early summer 2009, we had exhausted available data sources and most of our time. We suspended our research, reconvened our selection committee and began our final discussions. We reviewed all findings and began the painful process of deciding how to choose between two world-class manufactures and how to spend our funds in the most productive way. Our engineer crafted a decision-making table (see Table 2) to facilitate our process.

Table 2: Overhead Scanner Comparison
with specifications, attributes & operational issues

	Firm A	**Firm B**
Scanner Model		
Scanning bed size		
Overall working area physical footprint		
Camera CCD system		
Optical Resolution in ppi		
Speed at 600ppi color A1		
Start-up cost		
Ongoing annual cost – hardware and software maintenance contracts		
Delivery time		
Operational		
Test scan results and raw image processing		
Scanner Model		
Lighting design		
Ambient light control		
Computer interface		
Computer system		
Included software		
Customer service and company responsiveness		

This grid lists the decision points we were struggling with and how each scanner met our various challenges and expectations and includes variables such as scanning speed, ease of operation, customer service, and of course image quality.

Discussion

While this paper is not an evaluation of the firms in question, it is instructive for us to walk through some of their claims and our assessment of each firm's characteristics to share the flavor of our decision making process. Firm B's scanning speed was faster and their device, we thought, would be much easier for our student staff to use. The device from Firm A, however, had a reasonable scan time at the same resolutions and though more complicated to use, offered a wealth of camera and other adjustments designed to impact image quality and in the end, image quality and true fidelity to the original object was our main goal.

In our opinion, Firm A produced a slightly better digital image and most importantly, a raw image with minimal processing at the scan. Firm B's test scans produced noticeable noise and image distortion at their standard high speed setting and dealt with the problem by automatically adding image processing, which we had asked them to turn off at the test scan. As stated above, we felt, philosophically, that this was a very important point. If you intended to capture at very high resolution to gather archival quality images, then it should be done with true optical resolution. It made no sense to create pictures, which were not a product of photographic realism, but rather only a computer's vision of reality, performed to counter the limits of a camera.

We also discovered that Firm B was actually combining two images, in some way, to achieve an A1 scan at 600ppi. While Firm A's device scanned horizontally, the camera traveling the full width of the scan bed, the camera head in Firm B's device traveled vertically. This made the scan faster, but did not explain how all of the scanner bed or the object was scanned. Their device appeared to use two sets of image data, from two

CCDs mounted in the same camera. The scan data was then assembled to create the A1 digital image.

Furthermore, after doing a little math, we doubted that Firm B's 600ppi A1 image was based upon a true optical capture. The width of an A1 document is precisely 33.1 inches. Since Firm B's camera moved vertically, scanning the width line by line at 14,000 pixels, dividing the available 14,000 pixels by 33.1 gave a result very close to 400ppi.

In comparison, Firm A's camera traveled horizontally to scan the height line by line at 14,000 pixels. The height of an A1 document is 23.4 inches. Dividing the 14,000 camera pixels by 23.4 gave a result very close to 600ppi, making it a true A1 optical scanner.

Firm B was never able to fully explain to us the apparent contradictions in their product information. More than anything else, this clearly illustrates why comprehensive and exacting imaging standards are critical across the industry. Few scanner customers have the resources and expertise available to perform the months of research we carried out.

We signed the purchase agreement with Firm A in Mid-Summer 2009. Our biggest concern in buying their device was whether or not our student staff would be able to operate such a seemingly complicated device. The answer is a qualified "yes". What we have purchased is a very advanced scientific, photographic instrument. It demands very exacting camera and color balancing adjustments, which are performed by our engineer.

In our very practical model, however, the engineer sets the camera height, shutter speed, and aperture based upon the requirements of the scan job. After camera and color adjustments are in place, it is very easy for the student employee to manage the basic scanning operations.

We began the scanning of the Chicago Tribune pages in Mid-September 2009 and scanned all 5,000 pages by January 2010. The files are as huge as predicted and our Systems unit was forced to put a direct one gigabit pipe from DMC to our Library server room, just to handle our traffic. The images, even as reduced jpeg derivatives, are truly beautiful and can be seen at www.lib.msu.edu/branches/dmc/tribune/.

Conclusion

The choice of a capture device is the most important imaging decision a library or archive makes because it determines the maximum quality that can be achieved regardless of the other components of the digital library system. Given the importance of this activity, possessing the expertise to parse technical details and having workable schematics to compare vital variables cannot be overstated. Like most libraries when confronted with the task of investigating, selecting and installing digital library hardware, we performed our due diligence to make the best decision for our circumstances. The steps of the evaluation included: 1) defining needs and requirements; 2) investigating available solutions; 4) gathering data; 5) establishing criteria for rating; 6) applying weight to the criteria; and 6) arriving at a decision. This paper provides librarians with a basis on which to assess potential adoption decisions for a large-format scanner.

Epilogue

As we were completing our manuscript, we became aware of the publication of Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files (Federal Agencies Digitization Initiative, 2010). This document from the Federal Agencies Digitization Initiative in the United States provides a germane section entitled, "Quantifying Scanner/Digital Camera Performance."

References

Baird, H. (2007). "The state of the art of document image degradation modeling," Digital Document Processing, 261-279.

Boss, R.W. (2001), "Chapter 3: Image Capture," Library Technology Reports, 37(1), 20-28.

Burns, P (2007), "Ten Tips for Maintaining Digital Image Quality," Paper presented at the International Congress of Imaging Science, available at: http://www.imagescienceassociates.com/mm5/pubs/50Arch07BurnsWilliams.pdf (accessed October 20, 2010).

Burns, P. & Williams, D. (2002). "Improved Evaluation of Image Resolution for Digital Cameras and Scanners", Paper presented at the International Congress of Imaging Science, available at: http://www.losburns.com/imaging/pbpubs/35ICIS_sfr.pdf (accessed July 27, 2010).

Burns, P & Williams, D. (2002). "Refined Slanted-Edge Measurement from Practical Camera and Scanner Testing", Paper presented at the Image Processing, Image Quality, Image Capture Systems Conference, available at: http://www.losburns.com/imaging/pbpubs/33pics2002burnswilliams.pdf (accessed October 20, 2010).

Cornell University, "Moving Theory into Practice, Digital Imaging Tutorial," available at: http://www.library.cornell.edu/preservation/tutorial/conversion/conversion-04.html (accessed October 20, 2010).

Federal Agencies Digitization Initiative (2010), "Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files," available at: http://www.digitizationguidelines.gov/stillimages/documents/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf (accessed September 27, 2010).

Grimes, C., Riddell, L., & Nowson, C. (2009). "Consumer knowledge and attitudes to salt intake and labeled salt information," Appetite, 53(2), 189-194.

Image Science Associates, available at: http://www.imagescienceassociates.com/mm5/merchant.mvc?Screen=DONWIL&Store_Code=ISA001 (accessed October 20, 2010).

Jenkins, D. (2008). "How to Evaluate a Scanner," available at: http://www.iiri.com/resources/How%20to%20Evaluate%20a%20Scanner.pdf (accessed October 1, 2010)

Kenney, A., & Chapman, S. (1996). Digital imaging for libraries and archives: Dept. of Preservation and Conservation, Cornell University Library.

Lee, S. D. (2001). Digital imaging: a practical handbook New York: Neal-Schuman Publishers in association with Library Association Pub: pg. 50.

Library of Congress, "Guidelines for Electronic Preservation of Visual Materials," available at: http://www.loc.gov/preserv/guide/guide.html (accessed October 20, 2010).

Stelmach, M. (2006), "When Good Scanning Goes Bad: A Case for Enabling Statistical Process Control in Image Digitizing Workflows," Proc. Archiving Conf., IS&T, available at: http://www.imagescienceassociates.com/mm5/pubs/Good_to_Bad_Scanning.pdf (accessed October 20, 2010).

US Department of Agriculture and Department of Health and Human Services (2005), USDA. Dietary Guidelines for Americans, 6th ed., Washington, DC.

Williams, D. (2003), “Debunking Specsmanship,” DigiNews, 7(1), available at: http://worldcat.org:80/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file2003.html (accessed October 20, 2010).

Zeise, E., Williams, D., Burns, P. & Kress, W. (2007). “Scanners for Analytic Print Measurement - the devil in the details”, Paper presented at the SPIE, Image Quality and System Performance IV, available at: http://www.losburns.com/imaging/pbpubs/48EI07Zeise.pdf (accessed July 27, 2010).

Title: Hardware in libraries: making informed choices

Author Biography:

Shawn W. Nicholson
Michigan State University Library, East Lansing, MI USA
nicho147@msu.edu
has over a decade experience as a project manager and has written and lectured on digital curation activities. He presently holds administrative responsibility for the Digital Information Division of the Michigan State University Libraries.

Rick Peiffer
Michigan State University Library, East Lansing, MI USA
peiffer@msu.edu
is Information Technologist at the G. Robert Vincent Voice Library and the Digital and Multimedia Center with over 30 years experience creating unit best practices, regarding audio and video recording, playback and digitization.

John D. Shaw
Michigan State University Library, East Lansing, MI USA
shawj@msu.edu
is Assistant Head/Supervisor of the G. Robert Vincent Voice Library and the Digital and Multimedia Center. He was staff to the NSF funded National Gallery of the Spoken Word project and is presently project lead for the United Auto Workers Local 602/General Motors Oral History Project.