

# High Resolution Scanning for Preservation

## •Introduction

As an archaeology student by day and an archivist by night (well, actually afternoon) I have experienced first hand the challenges involved in the long-term preservation of our cultural heritage. Of primary concern is the functionally paradoxical dilemma of negotiating between the need to preserve and secure historically valuable materials, and the obligation to make such materials accessible to the public. In fact the point is not *just* to provide access to these valuable and often unique items but to *facilitate* that access.

Over the past decade or so, cultural resource managers have begun to look at digital imaging technologies as a way of resolving this dilemma. Indeed as these technologies continue to improve, the possibilities of a digital preservation agenda are fast becoming a reality. The advantages are clear: duplicates of text and visual materials archived in a stable medium; faster, wider distribution of those same files to the public, and the afforded opportunity to combine digital images with other media to create dynamic virtual displays. But with great power comes great responsibility (just ask Prometheus).

## •Significance

Even a relatively small digital project demands significant time and financial investment. Each step must be planned out well in advance and difficult choices have to be made at every juncture. A number of those choices will revolve around the selection

of appropriate technologies for the task; specifically scanning equipment and associated hardware and software. It is to this last point that I draw the reader's attention. The following discussion will focus on the technological implications of initiating a digital project to address preservation and/or archival needs. More specifically I will look at the role of high resolution scanning in this context and the factors which impact selection and utilization of this technology in a digital imaging project. I am paying particular attention to the scanning stage because in an archival context, decisions made at the point of digital capture ultimately affect the quality of images stored for posterity. Further, the quality of images generated from this stored copy, to be used in display or research, are likewise affected by decisions made at the point of capture.

## •Discussion

Digital scanning has become commonplace. As prices continue to fall, more and more folks are purchasing scanners for their own personal use. The images produced by these home-based units, however, are generally captured at a lower resolution (fewer pixels per inch) and as a consequence, are of a lower quality. In contrast, the images generated for an archival or preservation format tend to be of relatively high quality. There are several reasons for this. The digitization process, as stated previously, demands a significant investment in terms of time, money and resources. If you are going to do it, do it only once for each item. The underlying logic is, essentially, to capture the highest quality image possible given your technological and economic capacities, and then transfer the large files (resulting from high resolution settings) to a cheaper, high volume storage medium (i.e. CD-Rom, magnetic tape, etc..) (Besser and

Trant 2000). Compressed copies of that image can then be made for distribution or exhibit design purposes. A second reason for emphasizing high quality, high resolution scanning relates to the foundational impetus for digitizing in the first place: long-term preservation. In other words, if the original item is lost or damaged, its digital facsimile will be all that remains and should, by this virtue, attempt to approximate every nuance and quality of the original item.

### So, What is High Resolution Scanning?

The basic image capture system includes a scanner, software that can be used to manipulate and/or prepare image files for storage, and a storage medium (harddrive, CD, etc...). The scanner itself contains a lens, light source and a detector, the latter of which is usually identified as a charge coupled device (CCD). An item placed in the scanner is subjected to high intensity light and the resulting pattern is stored electronically on the detector as a *bitmap*.

#### **Helpful Web Sites Offering Review of Digital Imaging Basics:**

- Kodak Digital Learning Center. [Online]  
<http://www.kodak.com/US/en/digital/dlc/book3/>
- “Technical Recommendations for Digital Imaging Projects”;  
<http://www.columbia.edu/acis/dl/imagespec.html>

Put another way, the process of digitizing involves, through the use of this scanning or digital camera technology, the conversion of an item (photograph, text, drawing, etc..) into a digital image comprised of pixels or digital picture elements (Puglia 2000). Pixels are essentially little squares represented electronically by either single 1’s and 0’s or a series of 1’s and 0’s called bytes. *Resolution* refers to the number of

horizontal and vertical pixels used in the digital duplication of a particular image or object, and is discussed in terms of pixels per inch (ppi) or dots per inch (dpi). Generally speaking, the higher the number of pixels used per inch, the higher the resolution. Relatedly, the higher the resolution, the better image quality you will achieve.

*Optical resolution*, sometimes referred to as ‘true’ resolution, refers to the ppi settings inherent to the scanner being used; the resolution settings at the point of digital capture. This concept differs from *interpolated resolution* which involves the manipulation of a digital image through the use of software, after the scanning is completed. Though both steps are important to the digitization process, this discussion focuses primarily on the first scenario; conditions extant at the point of capture.

Another important factor conditioning the quality of digital images is the concept of *bit-depth* or *tonality*. “Bit-depth concerns the number of bits used to convey tonality for each pixel; that is, black and white, gray-scale, or color” (Technical Recommendations for Digital Imaging 1997). As a general rule, the higher concentration of bits per pixel, the larger the file size.

**1-bit or Bi-tonal**— two possible values, black or white; smallest resulting file size  
**8-bit or Gray Scale**—256 shades of gray ranging from pure white to pure black  
**24-bit Color**— tonal range of approx. 16 million colors; provide the greatest image detail and the largest file sizes

**High Resolution or Optical Scanning** then, relates to digital rendering of images through the use of scanning technology and /or digital cameras. Special attention must be paid to both ppi and bit-depth settings on the scanner in order to produce the highest quality images. A third parameter consideration is the overall dynamic range of the scanner. This particular concept refers to how wide a range of tones the device can

record (Besser and Trant 1995). These parameters must be set only after the nature of the source media (the items you will be scanning) has been assessed.

### A Few General Recommendations:

The following tables are summaries of general recommendations for resolution and bit-depth settings specific to media type. The first summary is from Reilly's 1996 recommendations to the Library of Congress. The second is from Steven Puglia's chapter on technical issues in the 2000 Handbook for Digital Projects (See References Cited).

(Summarized from Reilly 1996)

<b>Media Type</b>	<b>Conversion Method</b>	<b>Resolution</b>	<b>Archive File Format</b>
Black/White Text	flatbed or digital camera	1-bit, 600dpi	TIFF
Illustrations, Maps, Manuscripts, etc..	flatbed or digital camera	8-bit grayscale or 24-bit color, 200-300 dpi	TIFF
3-Dimensional Objects	digital camera	24-bit color, 200-300 dpi	TIFF
35 mm Black/White, Color slide or neg.	slide scanner or PhotoCD	24-bit, 2048x3072 dpi	PhotoCD or TIFF
Medium to large Format photo, slide Neg., transparency, Or color microfiche	drum scanner	24-bit, 4096x6144 dpi	TIFF
Black/White Micro-Film	microfilm scanner	1-bit 600 dpi	TIFF

(Summarized in Puglia 2000)

<b>Media Type</b>	<b>Resolution/Bit-depth</b>
Printed Type	1-bit, 600 ppi or 8-bit, 400 ppi
Textual Records	1-bit (200-600 ppi); 8-bit grayscale (200-400 ppi); 24-bit color, (200-300 ppi)
Photographs	8-bit grayscale (3000-5000 lines); 24-bit color (3000-5000 lines)
Maps/Plans/ Oversized	8-bit grayscale (200-300 ppi); 24-bit color (200-300 ppi)

The recommended settings listed above do not apply at all times. They are simply suggested ranges. Variation amongst medium types and quality of source materials demands that the project manager approach scanning on a case by case basis.

Generally speaking, the ideal would be to capture all selected materials at 24-bit depth and at least 600 ppi.

#### Summary of Factors Impacting Image Quality At Point of Capture

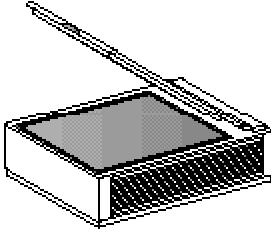
- The quality and nature of the selected source material(s)
- Scanning resolution, bit-depth settings and overall dynamic range of scanner
- Skill of scanner operator
- Particular type of scanning device used

The last item listed is important point. The following section will briefly cover some of the issues to be considered in the selection of an appropriate scanner for a project.

## Choosing the Appropriate Scanner for Your Project

There are several different types of scanners and the decision to choose a particular model(s) will depend on several factors. First, scanner types are generally limited to specific kinds of source mediums. For example, a slide scanner is dedicated to 35 mm slide film or other similar transparent materials. And, a flatbed scanner can usually only accommodate flat items of a particular size. So, before deciding on which technologies are appropriate for a given project, the nature of your source materials must be evaluated. What size and shape are the items? Are they of *different* sizes or shapes? Are the source materials in a fragile state? A second concern relates to the advantages, limitations and unique features inherent to individual scanner types. Finally, the budgetary allotment for new technology purchases and associated training will have to be considered.

## Scanner Types Relevant to the Archival Setting:



- *Main Types*

- **Flatbed**

*Description:*

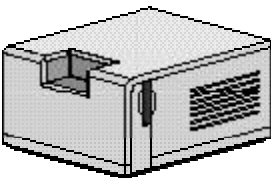
- The most typical of scanner types;
- Like a copying machine, an item is placed flat on glass surface and image is captured by diodes comprising the CCD apparatus below
- Usual resolution settings range from 300 to 1000 dpi

*Advantages:*

- Relative ease of use; attachment of sheet feeder allows processing of relatively large volumes of materials

*Disadvantages:*

- Limited to flat items, often of a particular size (8x10 or 11X17 in.)



- **Slide Scanner**

*Description:*

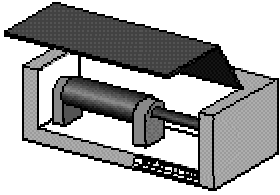
- Dedicated (medium specific) technology appropriate for 35 mm and transparent source materials; generally a small box with slot to insert slides

*Advantages:*

- Relative ease of use

*Disadvantages:*

- Limited to single medium



### •Drum

*Description:*

- One of the most powerful types
- Source material is wrapped around cylinder and rotated past a light source for image capture

*Advantages:*

- Renders high quality images and high resolution (typical resolution around 2700dpi)

*Disadvantages:*

- Cost is prohibitive
- Requires flat and flexible source material that can be wrapped around the drum
- Requires skilled operators

*Other Types:*

### •Video Digitizer

*Description:*

- Sometimes called a frame grabber

*Advantages:*

- Provides a fast scan time

*Disadvantages:*

- Image quality is limited by video image quality
- Limited to one medium

### •Copy Stand

*Description:*

- Essentially a digital camera on a stand capturing the image of an item likewise fixed to a stand

*Advantages:*

- Greater control over the positioning of the source material

*Disadvantages:*

- More time-involved

*(Note: Pictures of scanner types sampled from Besser and Trant 1995)*

Refer to these web-sites for a quick review of scanner types:

- Creating Digital Resources for the Visual Arts: Standards and Good Practice (**Section 3.5**) [http://vads.ahds.ac.uk/guides/creating\\_guide/contents.html](http://vads.ahds.ac.uk/guides/creating_guide/contents.html)

- Introduction to Imaging: Issues in Constructing an Image Database  
<http://www.getty.edu/research/institute/standards/introimages/>

## Pricing and Product Comparison

Other than mentioning budget as an important concern, I have not spent a great deal of time talking about specific costs associated with scanner selection. Generally speaking, prices can range from under \$100.00 to over \$20,000. There are also a wide variety of products on the market produced by a host of manufacturers. Again, planning is a big part of the budget process. The agency or institution should assess whether or not it is in the digital preservation game for the long haul. Products should be tested and compared before outright purchase. The risks are great but as technologies improve and prices continue to fall within an affordable range, the risks may become more manageable.

A thorough product comparison and analysis is beyond the scope of this discussion. The following two web references should offer some perspective. The first is a pair of links to Kodak product descriptions and the second links to an article in PCToday that compares various scanners and their features.

- KODAK**

- <http://www.kodak.com/US/en/digital/snapshot/scanners.shtml>

- <http://www.kodak.com/US/en/business/scanners/index.shtml>

- PCToday [On-Line] – An article comparing high resolution scanner products**

- [http:// www.pctoday.com/editorial/hth/970526.html](http://www.pctoday.com/editorial/hth/970526.html)

### **Summary of Factors Impacting Selection of Scanner Type**

- nature of the source materials
- advantages/limitations and features of individual scanner types
- budgetary allotment and overall costs of technology

### **In-House or Out-Source?**

This discussion has focused on considerations required for an in-house project. But one of the first decisions a digital project manager will have to make is whether or not digitizing should be completed in-house or contracted out to a vendor. The answer to this question will depend on the nature (type, physical state) of the materials selected for scanning, the volume of source materials, and the budget/resources available. There are advantages and disadvantages in both cases. An in-house approach will allow for greater control over all phases of the digitization process. Furthermore, the in-house project eliminates the need to transport fragile and often valuable items. The vendor on the other hand, can more easily handle large volume projects and, eliminates the need to spend inordinate amounts of time on the selection of appropriate technology and software. As a general rule in-house projects are more successful when the project is relatively small and the institution already maintains the necessary equipment (i.e. scanner/hardware/software) and trained staff (Puglia 2000).

## • **Summary**

This discussion, as promised, focused on the scanning phase of a digital imaging project. Particular attention was paid to the use of high resolution scanning in a preservation-oriented context. I looked at factors that can effect image quality at the point of digital capture as well as issues which impact selection of scanner types for particular project needs. Though associated software and storage mediums are also important digitizing components, their consideration in any depth would require an extended analysis. With regard to storage, prevailing wisdom contends that images produced for preservation should be saved in an uncompressed format (TIFF is the most common) and transferred to an economical storage medium such as CD-Rom that handle the large files. Duplicate images needed for display or research can be compressed into formats such as JPEG to facilitate distribution (Technological Recommendations for Digital Imaging Projects 1997).

To end, it was clear from my brief view of the landscape that with regard to establishing standards and guidelines for digital preservation, we still have a long way to go. The technologies involved are still relatively new and are evolving quickly. Preservation agencies are still trying to figure out how to pay for scanning technologies and incorporate them into their preservation agenda.

## References Cited:

### Besser, Howard and Jennifer Trant

1995 *Introduction to Imaging: Issues in Constructing an Image Database*. The Getty Art History Information Program.  
<http://www.getty.edu/research/institute/standards/intoimages/>

Conway, Paul

2000 Overview: Rationale for Digitization and Preservation. In *Handbook for Digital Projects: A Management Tool for Preservation and Access*, Chapter 2. Northeast Document Conservation Center, Andover, Mass.

Conway, Paul.

1996 *Preservation in the Digital World*. Washington, DC: Commission on Preservation and Access, March. <http://www.clir.org/cpa/reports/conway2/>

Frey, Franziska

1996 Digital Imaging for Photographic Collections: Foundations for Technical Standards. RLG DigiNews 1(3). <http://www.rlg.org/preserv/diginews/>

Grout, Catherine, Phill Purdy, and Janine Rymer

? Creating Digital Resources for the Visual Arts: Standards and Good Practice.  
[http://vads.ahds.ac.uk/guides/creating\\_guide/contents.html](http://vads.ahds.ac.uk/guides/creating_guide/contents.html)

Puglia, Steven

2000 Technical Primer. In *Handbook for Digital Projects: A Management Tool for Preservation and Access*, Chapter VI. Northeast Document Conservation Center, Andover, Mass.

## Useful Websites:

DIGITAL LIBRARY FOUNDATION

<http://www.diglib.org/preserve.htm>

CoOL (Conservation OnLine—Resources for Conservation Professionals)

<http://palimpsest.stanford.edu/bytopic/imaging>

*An Introduction to Digital Scanning*. Agfa Educational Publishing, Randolph, MA.  
[Online] <http://www.agfa.com/publications/>

Kenney, Anne and Oya Rieger. *Moving Theory Into Practice: Digital Imaging for Libraries and Archives*. Research Libraries Group, Mountain View, CA, 2000. [Online] <http://www.rlg.org/preserv/mtip-order.html>

Kodak Digital Learning Center. [Online] <http://www.kodak.com/US/en/digital/dlc/book3/>

*RLG DigiNews* on-line newsletter. [Online] <http://www.rlg.org/preserv/diginews/>

This paper is written by Sean Maroney  
for the course EDC385G Multimedia Authoring  
at the University of Texas - Austin