

Notes on Scanning from Fatima Mahdi, Hastings Historical Society

Getting Started

The first question to ask about digitization is not which scanner to buy, but **why** you are scanning. The use to which the digital files will be put will determine what you scan and how you scan, and even how you store the files when you are finished scanning. Do you want to have digital copies of every piece in your collection or only selected items? Which items, and what type are they? Do you intend to use your scanned images in your organization's publications? Will the publication be in color? Do you plan to use your images in a database? Will your users need to read the watermark on a letter or just the text? Will the image replace the original (for example, if you are scanning an item that has been loaned to you temporarily) or is it a backup or reference copy of an item that can be rescanned if necessary? Answering these questions will help you make better decisions. Answering them will also be important if you decide to apply for grants to digitize your collection.

Once you have decided what you are scanning and how the scan will be used, you will need to draw up some **guidelines** for the scanning process. Having guidelines will mean that you know what you have scanned, what format it is in, what the filename is, and where it is stored, without having to pull out a disk and look at it. Detailed guidelines will keep you from having to make the same decisions over and over again every time you turn on the scanner, and it will mean that everyone who scans items for the collection will be doing it in the same way. If you are concerned about color accuracy, consistent settings will help you evaluate your color balance.

Every scan will have to have a short unique **file name**. Many file formats give you the ability to attach longer comments to the file, but it is by these short names that you will organize and retrieve your digital files. Decide on a simple way to name files (accession number or catalog number are best), and a simple folder structure for grouping them together.

Spend some time **experimenting** with your scanner, scanning all the different kinds of items you might want to capture digitally (maps, photographs, brochures, pamphlets, calling cards, etc.) to see what kind of result you get for each type. You can even scan small three-dimensional items like campaign buttons or police badges. Print the scans, print them on a photocopier and an ink-jet printer, print them at 10 times the scanned size, zoom in on them and try to make out facial features in photographs, e-mail them, show them to people in your organization. Use them in all the ways you think you might want to use a scan. You may decide that you want to scan maps at difference settings than photographs. Then revise your guidelines. Decide what resolution you are going to use, what scale you will scan originals at, what file format you will use—and under what circumstances you might make exceptions to these rules.

Resolution

The resolution is the number of color samples the scanner will take within an inch. If you scan at 300 ppi, then the scanner is recording 300 separate spots of color in a single inch, arranged in a grid structure. The higher the scanning resolution, the better the quality of the image will be—up to a certain point. You may need to experiment to discover what that point is. The higher your resolution, the longer it will take images to scan. The file size will also be larger. The file will require more space to store, and it will take longer to open. Scanning at a very high resolution may be fine for a few items, but when you digitize a whole collection, storing and retrieving your scans can be a challenge. Most museums and archives advocate scanning items at 300 to 600 ppi.

However, the resolution at which you scan depends on how you are going to use the images, and what kind of detail the end user requires. At the moment you can use 100 ppi for your web site, but 600 ppi is needed to achieve reasonable quality for print production and 1200 ppi would be required for high-quality print production. For most items, 300 ppi will be sufficient. Some institutions use different settings depending on whether

they are scanning photographs or text. Most of today's scanners can scan up to 600 ppi, but you will need to check your scanner if you are interested in higher resolution.

Another aspect you need to consider is color. Scanning in color will result in a much larger file than if you scan in grayscale or black and white. You may decide to scan color items in color and use grayscale for newspaper articles. You may also have to decide what to do with sepia-toned photographs and maps with very little color. Are they worth scanning in color?

Attached you will find a table that shows the resolution choices of some major institutions. It was found at www.library.cornell.edu.

Size/Scale

Scanners also give you the option to increase the print size of your item. A 3 x 3 inch Polaroid scanned at 300 ppi will have less detail than if you increase the size (or scale) of your scan to 8 x 8 inches and then scan at 300 ppi. In looking for a standard for your guidelines, you can either decide to scan all your items at actual size, or enlarge them all to approximate the same size as you scan them. If you are concerned about using items from your collection in a publication like a newsletter, you might want to increase the size of very small items—or make a decision to scan everything at 300 ppi, except for items smaller than 3 x 3 inches, which you would scan at 600 ppi, giving you a higher-quality image that you can enlarge, if necessary. Scanner beds do not generally run larger than letter or legal size, so if you want digital images of larger objects, you might consider buying a digital camera and a tripod or copystand. The other alternative might be taking the item to a professional digital imaging company. The Westchester County Archives has equipment that can scan larger items and you might check with them to see if they would be willing to scan your items for a fee.

File Formats

The file format is the way the gray or colored dots in your scanned image are stored by the computer. Most museums and archives recommend saving images in TIFF (Tagged Image File Format), though PDF is increasingly popular for documents. Since many organizations favor using

TIFF and PDF, they will probably be supported by future versions of the software museums and archives are most likely to use. Other file formats, such as JPG, compress the image as it is stored. This means that the file size of the image is smaller, but some of the detail is lost. JPG, however, is also a very popular format, and more people are able to access JPG images on their computers and through e-mail. Again, it depends on the use to which you are going to put the scanned images. TIFF images can always be converted to JPGs when necessary.

For the debate about compression see: “Technical Feature: Lossy or Lossless? File Compression Strategies Discussion at ALA” (from the American Library Association meeting, 1999) at <http://www.rlg.org/preserv/diginews3-1.html#technical1>

Other Considerations

There are many settings on a scanner you need to experiment with—brightness, contrast and color balance. Be aware, however, that making adjustments for every image is time consuming. Differences in technology and settings mean that a scan will look different on your monitor than it does on someone else’s monitor, and it will look completely different when it comes off an ink-jet printer. Although you can open a scan and check to see what the resolution or size is, there is no way of telling how the brightness, contrast or color balance have been altered. Most scanners come equipped with excellent automatic settings for color and contrast. It is best to use these settings or find a good general setting to use for all your scanning and not spend time adjusting individual scans. It is useful to have a copy of Photoshop (or the less expensive Photoshop Elements) to make changes to your scans after you scan them. If your archival TIFF scans need to be cropped, rotated, color corrected for printing, and saved as a JPG for your organization’s next printing job, you can do it in Photoshop—when it is needed.

The following web sites and publications discuss digitization, scanning, and the standards various organizations have developed for their own collections:

The J. Paul Getty Trust's web site—"Introduction to Imaging"

http://www.getty.edu/research/conducting_research/standards/introimages/

Cornell University Library's web site—"Moving Theory into Practice: Digital Imaging Tutorial"

<http://www.cornell.edu/preservation/tutorial/contents.html>

Cornell University Library's hard copy book—*Moving Theory into Practice: Digital Imaging for Libraries and Archives* by Anne R. Kenney and Oya Y. Rieger (RLG, 2000)

The Research Libraries web site—"Guides to Quality in Visual Resource Imaging" which includes a good entry on "Planning an Imaging Project"

<http://www.rig.org/legacy/visguides>