

Digitization of Maps and Other Oversize Documents

Janet Gertz Columbia University Libraries

Paper maps (and other oversize documents such as architectural drawings) contain a wealth of fine details composed of graphic and textual elements. They include:

- The drawing of the location
- The use of graphic entities like elevation lines or symbols for cities of different sizes
- Printed names of countries and other features
- Color, which carries information through varied patterns and intensities.

When they are large, maps with fine details present special difficulties for digitization. There can be a huge disparity between the size of the document and the size of the smallest meaningful element that must be made visible online or in printouts. Fine detail requires high resolution scanning, and the result is very large file size. File manipulation, storage, delivery, and display all become much more complicated.

Even the mechanics of scanning are affected. Many flatbed scanners have size limitations and cannot handle large maps. Scanning may require film intermediaries such as 4x5 transparencies or single-frame microfiche, where the original object fills the body of the microfiche. Thirty-five mm slides are too small to fully capture details on large maps. When originals are not only oversized but also brittle, working from a film intermediary will put less strain on the fragile original. Some loss of quality will result because the film version is one generation removed from the original. However, fully legible images can be produced from film intermediaries, always, of course, given that the transparency or microfiche is itself carefully made and then scanned with sufficient resolution and appropriate tonality.

Scanning Parameters

Determining capture parameters follows the same rules as for other documents.

- Decide on the appropriate tonality, usually gray-scale or color. Color on most printed maps is important as a coding device, not for its precise hue as it is in art works. Nevertheless, a standard color bar should be included during scanning even when sophisticated color management is not a requirement. (For a discussion, see Ester, 1996.)
- Identify the smallest meaningful element, often a thin line.
- Determine how many pixels are needed to capture the smallest element legibly.
- Calculate the necessary resolution, often 200-300 dots per inch when scanning in 24-bit color. For a discussion of resolution and related issues, see Gertz *et al.* (1996) and Allen (1998).
- Whether the map is digitized directly or through a film intermediary, always include a ruler in the image so that dimensions and distances are unambiguous. As an example, consider a hypothetical map two feet across and three feet long.

- The smallest textual elements are numbers less than 1 mm high that record elevations.
- The smallest meaningful elements are the thin lines used to indicate elevation.
- Ten different colors serve as codes, patterned as dots, parallel vertical and horizontal lines, and other graphic devices. A scanned version of acceptable quality would permit users working on screen and with printouts to:
 - Read the 1 mm text
 - See unbroken elevation lines
 - Clearly distinguish all color code patterns.

Assume a minimum of 200 dpi and 24-bit color is needed to achieve legibility of the 1-mm text and the lines and code patterns. For a map 36" wide, 200 dpi multiplies out to 7,200 dots across the surface of the map. If a film intermediary is used, then the effective resolution must be calculated as well. Effective resolution refers to resolution relative to the size of the original document. A transparency still requires 7,200 dots across the map to capture the same degree of detail. The map on the transparency is perhaps only 4" wide. It must be scanned at 1,800 dpi to get the same level of detail.

To calculate the file size, use the formula given in Kenney and Chapman (1996), p. 20. The product of digitizing oversize documents is clearly a series of very large files. This has implications for the image creator in terms of storage, retrieval, and display. Large files take up a great deal of storage space. Enough memory must be available for images to be loaded and manipulated. Backing up files, creating derivatives, and transmitting files absorb a significant amount of time and storage formula: (height x width x bit depth x dpi²) / 8 original map: (24" x 36" x 24 x 200²) / 8 = 103,680,000 transparency: (2.667" x 4" x 24 x 1800²) / 8 = 103,680,000 media.

Problems of Access to Scanned Large Maps

The nature of these files also translates into problems for users trying to access and navigate within digital images of large maps.

- The high-resolution image in which all of the details are visible is too big for users to access or manipulate easily, given current delivery mechanisms and the capacity of common computers.
- When derivatives of the original high-resolution files are provided for access, they are often JPEG versions with considerably reduced resolution. If the resolution is low enough to make files easy to access, the finer details in the images may become illegible.
- Only part of the map image fits on screen at one time. When using the paper document, readers orient themselves to salient features through peripheral vision while focusing closely on details. On screen, it is easy to become disoriented because most of the image is not visible.
- With a paper map, it takes a single glance to follow features such as roads or boundaries from one edge to the other, but on screen it takes continued scrolling. Comparing widely separated details becomes awkward at best if they are not visible simultaneously.

Benefits of Scanning Large Maps

Despite these difficulties, there are a variety of ways to benefit from scanning large maps.

- Use the high-resolution images to produce high quality printouts to replace brittle originals.
- Derive lower resolution versions from the high-resolution master images to serve as reference-quality images and reduce unnecessary handling of brittle originals.
- Put the images on CDs and view them directly rather than trying to deliver them over a network. The workstation must be capable of handling the large files.
- Scan large maps in sections to generate a group of high-resolution files of manageable size. This entails use of software packages for managing the separate files and concatenating them as the user moves from one to the next. It also can complicate the creation of high-resolution printouts.
- Investigate some of the new compression software that permits the user to access a lower resolution image and then zoom into higher resolutions without manipulating the whole high-resolution file locally. One such product is Lizardtech's Multi-Resolution Seamless Image Database (see <http://www.lizardtech.com>); a number of other packages are available.

In Conclusion

- The size of the original, in proportion to the size of the smallest meaningful element, determines the needed resolution.
- File size governs the ability to store, retrieve, and display an image.
- Excellent images will fail to satisfy users if they cannot be accessed or if equipment and software are not well suited to working with large images. Speed and smoothness of scrolling and zooming are important.
- Planning the user interface must be part of initial project design. To view a selection of approaches to scanned maps, see:
- American Memory project:
 - <http://memory.loc.gov/ammem/gmdhtml/gmddigit.html>
- British Columbia Archives and Records Service:
 - <http://www.bcars.gs.gov.bc.ca/cartogr/general/maps/html>
- Library of Virginia: <http://image.vtls.com/BPW>
- University of Connecticut: <http://magic.lib.uconn.edu/magic/exhibits/>
- Atlantic Neptune: <http://mercator.cogs.nscs.ns.ca/neptune.html>
- National Oceanic and Atmospheric Administration:
 - <http://chartmaker.ned.noaa.gov/ocs/text/MAP-COLL.htm/>
- David Rumsey Associates: <http://www.davidramsey.com>

Sources

Allen, David. "Creating and Distributing High Resolution Cartographic Images," *RLG DigiNews* 2:4, 1998. <http://lyra.rlg.org/preserv/diginews/diginews2-4.html#feature>

Ester, Michael. *Digital Image Collections: Issues and Practice*. Washington, DC:

Commission on Preservation and Access, 1996.

Gertz, Janet, Robert Cartolano, and Susan Klimley. *Oversize Color Images Project*, Columbia University, 1996. <http://www.columbia.edu/dlc/nysmb/>

Kenney, Anne, and Stephen Chapman. *Digital Imaging for Libraries and Archives*.

Ithaca, NY: Department of Preservation and Conservation, Cornell University Library, 1996.