



Conserve O Gram

September 2000

Number 19/21

Planning Digital Projects For Preservation And Access

Digital technology offers many advantages, such as image enhancement, copy fidelity, multi-user access, and many output formats. The biggest advantage is improved access. Digital files can be shared over the Internet with many millions of viewers.

Preserving digital files involves managing an ever-changing infrastructure of hardware, software, file formats, and storage media. Major changes take place in hardware every 1-2 years and in software every 1-5 years. Proprietary file formats are supported only briefly and may vary from vendor to vendor. The storage media (diskettes, tape, compact disks, DVDs, zip, and jazz drives) last 10-200 years, but the hardware and software to play them vanishes, changes, and/or self-destructs in < 5-15 years.

Fragile digital files require regular refreshing (rewinding and recopying) and vigilant migration (translation to new software and hardware). When refreshing and migration costs are added to the cost of maintaining software and hardware, the cost of maintaining digital files over time becomes many times that of keeping paper.

Use digital media (optical or magnetic storage) for access, not to permanently replace analog documents (such as microfilm, paper, audio and videotapes, and film).

Basic Concepts: Read about well managed digital projects on the Web, as well as in books. See the **References**. Learn these key concepts:

- **Compression:** the process of forcing more data into less space to speed processing, storage, and transmission. Compression may be *lossless* (less compression, but no data

loss), or *lossy* (deep compression with data loss). Don't expect lossy file formats (such as JPEG files) to look the same after compression.

- **Dynamic range or bit depth:** the number of colors represented, such as 8-bit images (allows up to 265 colors) or 24-bit images (allows up to 16 million colors).
- **Image enhancement:** the use of scanning/editing software to correct/edit an image.
- **Pixel:** a picture element or tonal value in binary code.
- **Resolution:** the image's clarity/definition (often listed as height by width in pixels, [50 X 100 pixels] or as dots per inch [200 dpi]).
- **Threshold:** the bitonal scanning setting at which gray is read as black or white.

As you begin a project, know your:

- **Mission:** What is your institution's mission, scope of collections, access policy, and the requirements of your funding source? Your project should support these goals.
- **Audience:** Decide who your digital project will serve in terms of age, education, socio-economic level, and discipline or topical focus. What does your audience need?
- **Priorities:** What is the importance of speed, cost, image quality, and quantity?
- **Functional goal:** Will you use your digital files for reference, exhibits, publication, and/or on the Web?
- **System capabilities:** Do you need to upgrade your system memory, processor speed,

connectivity, band-width, monitor, scanner, and storage device? What is the status of your storage, computer infrastructure, and support? How long will it take to upgrade?

- **Staff expertise:** Do they have skills in contracting, scanning, and managing infrastructure and migration? If not, can you contract or hire new staff?
- **Content:** Before you determine your scan size/quality, select materials and decide what aspects and attributes of your original you must capture. Are the candidate items to be scanned authentic (unaltered) and in good physical condition? If not, you must stabilize them before scanning or scan from a copy.
- **Content documentation:** Review all captions, catalogs, and collections management information. Are they complete, accurate, and appropriate?
- **Legal rights and restrictions:** Examine the legal status of the collections (copyrights, privacy/publicity rights, defamation issues, and obscenity issues) and negotiate with rights holders to acquire the right to reproduce an item *before* you digitize.
- **Collection sensitivities:** Evaluate your collections for donor restrictions and sensitivities (such as sacred site, burial, and endangered resource locations). Consult with all major stakeholders who have interests in how the collection is made accessible, including donors, creators, researchers, staff, and subjects of the materials to ensure that no damage results from digital publications.
- **Collection stability and staff time:** Making images accessible digitally increases reference and duplication demand for the originals, as well as staff workload. Ensure that you can handle added work before you mount copies on the Web.

Getting Started: Start with a small and easy project in a standard format that allows you to gain expertise. Select a widely compatible or

“open system” architecture if you are involved in planning the computer infrastructure. Produce large master files in non-proprietary file formats (such as TIFF) that have lossless compression. Store your master files offline in an uncompressed or lossless compression format. Back-up, refresh, and migrate all files regularly. Produce derivatives (usage copies), such as thumbnail images for online or Web use in common file formats such as TIFF or GIF. Maintain excellent metadata (see below). Calibrate and test your equipment regularly.

Contractors: Don’t select contractors based solely upon the recommendations of other vendors or friends. Choose contractors based upon their experience with similar materials, references, range of scanning capabilities, the contract terms, and their relationships with other potential partners. Ensure that your contract states that the park receives all copyrights for new images as well as all products (negatives, copies).

Selection: Ensure that what you digitize fits your organizational mission and collecting policy. The cost of creating a digital file over several decades is only a small percentage of the cost of managing the file, so ensure the file has continuing value to the park and program.

What Are the Steps in Digital Work?

- Determine the project audience, purpose, and goals and match them to the institutional mission and collecting goals.
- Assess the candidate items for scanning:
 - Does the collection have condition and treatment needs?
 - Are there problems with mats/frames/housings?
 - Is cataloging/indexing at the appropriate level?
 - Can the original object’s content be fully captured?
 - Are there legal, cultural, ethical, or donor restrictions?

- Select the items to be digitized.
- Retrieve and stabilize the items.
- Benchmark your digital image using the formula for predicting image quality, file size, storage costs, and access time. See Kenney and Chapman (1996).
- Prepare documentation of originals (preliminary numbering, indexing, cataloging, and description of originals).
- Rehouse and pack the items for transport.
- Transport the items to the contractor or scanner/camera and produce photographic or microfilm copies for preservation purposes.
- Scan the photograph/microfilm copies *or* scan the original. Record scanner settings.
- Format the file. Prepare copies in alternate sizes/formats (derivatives).
- Label (name/number) and record metadata (see below) on digital master (first digital surrogate) and on derivatives (later digital copies produced for use).
- Pack the originals and digital surrogate files for transport.
- Transport the originals and digital surrogates to the repository.
- Unpack and rehouse the original items.
- Manage the digital file by backing it up, storing it, checking it regularly, correcting errors that occur over time, refreshing it, and migrating it as hardware/software change. If this is not done, the file will be lost.
- Manage the network by upgrading and repairing it as necessary.

Documentation and Quality Control

Documentation and quality control is necessary to ensure useful digital records. It consists of several steps, including:

- metadata (see below)

- quality control check of files (see below)
- refiling of the original objects
- integration of images with text
- cataloging and cross-referencing
- verification of work completion
- contracting paperwork

Metadata: Document the authority, contents, and format/structure of what you scan, including:

- ***management data***, such as image specifications, item preservation, and use
- ***provenance data***, such as the history of creation of the digital and source file
- ***rights data***, including copyrights and access conditions, model releases, and ownership
- ***organizational data***, including tables of contents or targets (bibliographic labels and use instructions). See Baca (1998).

To maintain quality: Maintain a consistent color space and system by selecting appropriate hardware and software. Use color/tonal bars and targets during all scanning and when calibrating monitors/printers so that colors are equivalent. When picking your file format/size, match the scan size to the important aspects of the original. Learn how file formats will affect image quality.

Benchmarks: Develop quality models and standards to systematically use as management tools for predicting the probable results of imaging techniques, including:

- the height of the smallest significant character
- the desired image quality
- the scanner's resolution capabilities

Time and Cost Estimates for Digital Work

Know the quality level required, based upon:

- the long- and short-term plans for the scan

- the original's format, type, and amount of information
- what you wish to capture from the original
- the quantity/variety of document types to be scanned

Factor in the hardware, software, and staff necessary to do all stages of work, including refreshing and migration. Use these estimates:

- The original scan of your item is 32% of the cost of the total digital conversion.
- Cataloging, description, and indexing is 29% of the cost of the total digital conversion.
- Quality control and file management is 39% of the cost of the total digital conversion.

An average digital image conversion runs between \$17.65-\$23.25, according to a recent National Archives study. Time and cost estimates won't remain stable throughout the project. As you gain experience you will gain speed and improve quality. See Puglia (1998).

Cost estimates for maintenance of digital files for 10 years range from 14%-60% of the initial conversion cost. Service Bureaus are an option for maintaining digital files, although they charge between 275-796% of the initial conversion cost to maintain files for a decade.

As Steve Puglia notes in *The Costs of Digital Image Projects*, "Institutions will only be able to justify the maintenance of digital images that are used...[and] will need to perform a cost-benefit analysis to determine how much use is necessary [to keep them]. Analog storage, such as keeping the originals or microfilm, offers greater longevity and less risk for long term retention. Hybrid approaches [digital for access plus film for preservation] ...are likely to continue to be the most viable."

The *Conserve O Gram* series is published as a reference on collections management and curatorial issues. Mention of a product, a manufacturer, or a supplier by name in this publication does not constitute an endorsement of that product or supplier by the National Park Service. Sources named are not all inclusive. It is suggested that readers also seek alternative product and vendor information in order to assess the full range of available supplies and equipment.

References

- Besser, Howard, and Jennifer Trant. *Introduction to Imaging*. Santa Monica, CA: Getty Art History Information Program, 1995. On the web at < www.getty.edu/gri/standard/introimages/index.html > .
- Baca, Murtha. *Introduction to Metadata: Pathways to Digital Information*. Santa Monica: Getty Information Institute, 1998. On the web at < www.getty.edu/gri/standard/intrometa/ > .
- Kenney, Anne, and Stephen Chapman. *Digital Imaging for Libraries and Archives*. Ithaca, NY: Cornell University, 1996.
- Ostrow, Stephen E. *Digitizing Historical Pictorial Collections for the Internet*. Washington, DC: Council on Library and Information Resources (CLIR), 1998. On the web at < www.clir.org/pubs/reports/ostrow/pub71.html > .
- Puglia, Steven. *Creating Permanent, Durable Information*. Washington, DC: National Archives and Records Administration, 1998. Also see *The Costs of Digital Imaging Projects*. On the web at < www.rlg.org/preserv/diginews/diginews3-5.html > .
- Research Library Group Preservation Program. *RLG Worksheet for Estimating Digital Reformatting Cost*. On the web at < www.rlg.org/preserv/RLGWorksheet.pdf > .
- Zorich, Diane M. *Introduction to Managing Digital Assets: Options for Cultural and Educational Organizations*. Santa Monica: The Getty Information Institute, 1999.
- Diane Vogt-O'Connor
Senior Archivist
Museum Management Program
National Park Service
Washington, DC 20240

The series is distributed to all NPS units and is available to non-NPS institutions and interested individuals by subscription through the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402; FAX (202) 512-2250. For further information and guidance concerning any of the topics or procedures addressed in the series, contact NPS Museum Management Program, 1849 C Street NW (NC 230), Washington, DC 20240; (202) 343-8142.