Library of Virginia

DIGITAL IMAGING GUIDELINES

September 2008

URLs and cites updated Aug. 2018

Scope

This guideline is applicable to all state agencies, public institutions of higher education and local government entities. This guideline is intended to apply to imaging systems that involve routine systematic processing, storage, or retrieval of documents, pictures, maps, drawings, and similar items that constitute public records. While the recommendations set forth in this document reflect current best practices in the field of digital imaging, this set of guidelines is not meant to define mandatory standards.

Legal Framework

The *Code of Virginia* citations below address important controls on public records in the areas of records management, electronic signatures, privacy, evidence, etc. This is not a comprehensive list of imaging-related legal requirements.

- Virginia Public Records Act (*Code of Virginia*, § 42.1-76–§ 42.1-91)
- Virginia Uniform Electronic Transactions Act (*Code of Virginia*, § 59.1-479–§ 59.1-498)
- Copies of Originals as Evidence (*Code of Virginia*, § 8.01-391)
- Virginia Freedom of Information Act (*Code of Virginia*, § 2.2-3700–§ 2.2-3714)
- Government Data Collection and Dissemination Practices Act (*Code of Virginia*, § 2.2-3800–§ 2.2-3809)
- Virginia Civil Remedies and Procedure (*Code of Virginia*, § 8.01)

Purpose

The purpose of the following guidelines is to provide best practices for public bodies that are investigating the use of imaging systems for storage and retrieval of public records. Agencies and localities are responsible for implementing appropriate policies, procedures and business practices in order to ensure that an imaging system protects the authenticity, reliability, integrity and usability of public records. As such, the guidelines are designed to identify critical issues for public officials to consider when designing, selecting, implementing, operating and maintaining imaging technology. The guidelines are divided into four sections with two appendices:

Introduction

Imaging, as used in these guidelines, refers to the capability to capture, store, retrieve, display, process, manipulate and distribute a digital representation of a document. Micrographics refers to the techniques associated with the production and handling of microfilm, microfiche and related storage technologies based on retaining a photographic representation on film. Micrographics are sometimes considered imaging applications because they have many common features. However, unlike imaging, micrographics are not digitally stored and do not have the transmission and manipulation flexibility of digital images. In some systems, imaging and micrographics technologies are combined in a hybrid system.

A digital image is a representation of a two-dimensional image as a finite set of digital values, called picture elements or pixels. At the most basic level, a digital image is a matrix of pixels digitally recorded as a series of 0's and 1's. Although an image is a matrix of individual pixels, these pixels are densely concentrated at rates of several hundred to several thousand per inch. When viewed, they appear to the eye to have the pattern and shape of the original document, picture, map, drawing, X-ray, etc.

Unlike other types of coded and digitized computer data, imaged data carries no intrinsic meaning but simply creates a picture that must be interpreted by the viewer to have significance. For example, the letter "A" in a word processing document is a digitized code that can be recognized as the letter "A" by other computer application programs at other times, using other hardware platforms. In an image, however, the letter "A" is not coded to be recognized by an application program, including the one managing the imaging display. Only the viewer can recognize and give meaning to the letter shown on the screen. This distinction is important in understanding the relationship of imaging technology to other information technologies, in particular for designing indexing strategies and evaluating the proper role of data capture techniques such as barcode recognition (BCR) or optical character recognition (OCR).

Per the Copies of Originals as Evidence section of the *Code of Virginia* (8.01-391), public records may be imaged, or reformatted, and maintained electronically. Furthermore, the imaged copy may be used as the record copy and the paper original can be destroyed. Public officials are responsible for ensuring that the imaged records are accurate copies of the originals and must provide access to them for their entire retention period. In addition, there must be a quality control process in place to certify that the imaged records are visually inspected for legibility and integrity, as well as an indexing system to allow for easy access and retrieval. If the records are considered vital a security copy of the images, indexing system and software application must be stored off-site.

Please consult with your Records Management Analyst at the Library of Virginia (LVA) prior to reformatting permanent records. Original records that have a permanent retention period must be offered to the LVA prior to being destroyed.

DIGITAL IMAGING GUIDELINES

SECTION 1: PROJECT PLANNING

The key to the successful design, integration and implementation of a digital imaging system is thorough planning. There are a number of planning tools available to state agencies and localities through the Virginia Information Technologies Agency (VITA). The Commonwealth Project Management (CPM) section of the VITA Web site provides links to established policies, standards, and guidelines for the management of Information Technology (IT) projects in the Commonwealth. Additionally, the CPM site provides tools and templates to assist project sponsors, program managers and project managers in project initiation, planning, execution, control and closeout. The documents and templates provided are consistent with best practices established by the Project Management Institute.

Before the decision to carry out an imaging project is made, a project scope should be established and a feasibility study should be conducted. The study will not only determine whether a project should be undertaken but also whether the project will be completed in-house, by an outside vendor, or using a combination of the two approaches.

At a minimum, the feasibility study should consist of the following components:

1. **Institutional Support**

   A successful imaging initiative hinges on upper management support. The implementation of a digital imaging system may require complete reevaluation of workflows, redefinition of staff expertise and reallocation of funding. As such, support must come from the “top down” in order for staff to understand and embrace the complexity of such an initiative and its importance to institutional missions and priorities.

   Upper management must also provide the necessary direction, identifying the appropriate staff to undertake each step of the process. Only with this support will staff effectively participate and willingly learn new skills and workflows in order to operate, use and maintain an imaging system.

   Imaging projects require a combination of skills from a variety of staff with different areas of expertise. The project team should be led by a project manager who meets the qualifications and training standards established by VITA’s Project Manager Selection

and Training Standard. The project team should also include members from the following departments:

- Information Technology
- Records Management (in most cases, the records officer)
- Finance
- Facilities
- Human Resources
- Legal

2. Needs Assessment

Implementing a digital imaging system significantly impacts the current work processes by requiring personnel to create, retrieve, share, use and store documents in different ways. A needs assessment must include both records and workflow analyses.

A. Records Analysis

A records analysis assesses existing operations to determine which records—based on content, legal requirements, use levels, or work processes—are best suited for digital imaging systems. Imaging is best suited in situations where:

- Volume of documents is high
- Documents require action to provide a government service, as opposed to being primarily stored in files, libraries, or archives
- Documents are moved through multiple workstations and/or accessed multiple times during processing
- Routing and work processing for documents are stable and predictable
- Retrieval access time is especially imperative
- Control of the document during processing is important (e.g., tracking documents through the work process; avoiding the physical loss of a document; monitoring workload distribution, backlogs, and processing productivity; and ensuring that the document will be re-filed properly)
- Documents are intellectually, but not physically, associated (i.e., records on a similar subject are likely to be retained in logical electronic “file folders” after scanning, which may both speed up retrieval time and promote a better understanding of the records as a whole)
- Document enhancement is beneficial (i.e., for improved legibility)

B. Workflow Analysis

A workflow analysis assesses the processes of records creation, access and retrieval to determine areas where reengineering can improve operational efficiency. The workflow analysis should consider the following questions:

- Once a document is created or received by your office, what does your office do with it? Who sees it? Where is it filed?
- Does the document require any official signatures or approvals?

Once these questions have been answered, consider the existing workflows in terms of the changes that will be required due to imaging:

- How will that same document be handled once your imaging system is in place?
- What safeguards are in place to ensure that proper approvals have been received?
- If a file is shared between departments, who will be the manager of the document?
- Who will image the document?
- What types of quality controls will be established?
- Will any information need to be redacted or pages restricted from public view?
- How will the document's retention and disposition be managed in a new format? (see Section 3: System Implementation and Section 4: Archiving and Long-Term Maintenance)
- If you intend to destroy the paper version after imaging, how will that destruction be handled?
- Will the document remain filed for a period of time before destruction?

3. Cost Benefit Analysis

Digital imaging should make financial or business sense for your agency or locality. Justifying the cost of a digital imaging system involves a financial comparison between current and proposed record-keeping systems to help make a procurement decision. The cost-justification goal of a digital imaging system is to offset costs with cost savings or business benefits. For example, an agency or locality may offset the cost of the equipment, software or outsourcing fees by reducing personnel and storage costs or

allowing the existing staff to process work more efficiently through the improvement of work processes. A business benefit may justify a cost that is not offset. For example, public access to records may be greatly improved through the implementation of a digital imaging system.

A typical cost justification includes the following:

- A study of current operations
- Potential improvements to current operations (e.g., better customer support, improved efficiency, etc.)
- Proposed system architecture
- Equipment pricing
- Financial measures including payback period, rate of return and net present value

To determine the benefit derived from the new system when compared to an existing paper-based system, an agency or locality may consider the following existing costs:

- File creation including file folders, labels, paper, file tracking system and labor to create files and to add to system
- File maintenance including filing equipment, floor space to store and access files, labor to retrieve, copy and re-file documents, time spent waiting for information retrieval, cost of misfiles and cost of lost files
- File disposition including boxes for off-site records center storage, labor to move files from active to inactive storage, off-site storage and destruction (recycling, pulping or shredding)

Initial investment in equipment, staff training, capture and conversion, handling, storing, and housing originals, producing derivative images, cataloging and building the image database system and developing Web interfaces are all possible areas of cost for an imaging system. In addition, ongoing costs of maintaining data and systems over time include storage media, maintenance contracts, media refreshment, technology upgrades, hardware replacement, migration, training and labor. Also keep in mind that capture and conversion of data often comprise only one-third of the total costs, while cataloging, description and indexing comprise two-thirds of the total costs.[1]

[1] Western States Digital Standards Group Digital Imaging Working Group. Western States Digital Imaging Best Practices, Version 2.0. University of Denver and the Colorado Digitization Program; Denver, 2008, p. 6.

A Cost Benefit Analysis Spreadsheet Modeling Tool is available through VITA's Project Management Web site. The modeling tool instructions and template can assist project managers as well as agency and locality management to determine the best solutions for business problems.

For additional guidance on budgeting activities involved in imaging, see the Research Libraries Group (RLG) *Worksheet for Estimating Digital Reformatting Costs*.

4. Alternatives Assessment

Imaging and micrographics each yield different, yet excellent, benefits. For storage and retrieval applications, relatively low storage volume and high retrieval rates favor imaging, while high-volume, low-retrieval applications favor micrographics. When comparing alternatives, be certain to consider ongoing costs of access over time. You may decide that a combination approach—scan to microfilm—may be the best solution for your agency. The following comparison lists some features of the two technologies:

IMAGING	MICROGRAPHICS
Poor/mixed standards	Developed standards
Access via telecommunications to geographically dispersed locations	Distribution by mail to geographically dispersed locations
Retrieval speeds in seconds to minutes depending on configuration	Automated indexing gives reasonable retrieval speeds
Already digitized	Can be scanned and digitized
Available to multiple users in multiple locations within a workplace	Available to one user at a time in a workplace
Workstations combine images, database access, and other functions	Single-purpose workstation
Resolution is available to high levels	Resolution is adequate depending on source document
Strong in file folder applications	Weak in file folder applications
Image can be enhanced for viewing	Image as filmed
Repeated viewing does not degrade image, additional copies do not lose quality	Repeated viewing degrades media, recopying reduces image quality
Good backup, good security	Good backup, good security

5. Technology Analysis

The rapid rate of technological change must be considered and built into your plan. First, a technology analysis must verify that reliable hardware, software and storage media can be acquired and will integrate with your existing systems. Second, the analysis must verify the ability to migrate records to newer technologies as necessary. Finally, the analysis will confirm whether or not in-house staff members have the proper skill sets and training to manage a digital repository.

Because most agencies and localities do not have the appropriate scanning equipment, software or staff expertise to execute a large imaging project, vendors have become integral to the world of digital imaging. Quality varies among vendors, so selecting the right one is crucial to your project. Vendors provide imaging services, technical advice, and sometimes long-term maintenance of the resulting images. To make comparisons and validate vendor claims, it's always best to get live demonstrations using your materials. Becoming familiar with digitizing technology and the terms used by the industry will help in your vendor selection (See also: Section 2: Technical Specifications and Selection).

6. Project Scope

Finally, whether the project will be completed in-house, by an outside vendor or using a combination of the two approaches, you should have a clear understanding of the scope of your project by addressing the following questions:

- How much material will be imaged? Over what time period? Is back-file conversion necessary or is “day forward” imaging adequate?
- What types and sizes of material will be imaged? Textual documents? Photographs? Maps?
- Who is the intended audience? Staff members? Researchers? The general public?
- What is the desired end product? A document management system or electronic records management system? A searchable online collection?
- How and where will the audience access the images? How frequently?
- How will disposition be handled at the end of the required retention period?

DIGITAL IMAGING GUIDELINES

SECTION 2: TECHNICAL SPECIFICATIONS AND SELECTION

System Hardware and Software

Imaging systems should incorporate a standards-based environment. Such an environment may include a comprehensive set of standards-based interfaces, services and supported formats that enable portability and interoperability of applications and data. Standards allow systems to change over time without significant risk of losing records. Standards also facilitate the importing and exporting of images.

All aspects of the design and use of the imaging system should be documented. Documentation should include:

- Administrative procedures for imaging, storing and retrieving records
- Technical specifications for all components of the system
- Problems encountered in the operation of the system and measures taken to address such problems
- Hardware and software modifications

For additional information regarding documentation requirements, see the Virginia Information Technologies Agency (VITA) Information Technology Resource Management (ITRM) Policies, Standards and Guidelines.

If business needs dictate, it is recommended that digital images be captured into an integrated Electronic Document Management (EDM) system and Electronic Records Management (ERM) system, also known as an Enterprise Content Management (ECM) system. ECM is the technology that allows organizations to manage their unstructured content, which includes “born digital” documents (e.g. Word, Excel, PDF), media files (audio and visual), and scanned images of paper documents. Benefits of an integrated EDM/ERM or ECM system include the ability to:

- Prove the authenticity, reliability and integrity of records
- Provide audit logs of access history
- Utilize workflow functions to track the processes of creating and distributing records
- Manage records retention and disposition (See also: Section 3: System Implementation and Section 4: Archiving and Long-Term Maintenance)

In 2007, IBM FileNet P8 was selected as the Commonwealth standard for ECM software. Services available include business process management, email archiving, e-forms, records management, storage connectors, business activity monitors, Web site templates, redaction software and more. This business process software achieves the objectives set out in Governor Kaine's "Paperless Government" Initiative through improved digital capture, document storage and retrieval services. A Shared Services platform also allows smaller to medium-sized agencies that cannot justify their own ECM solution or larger agencies wishing to pilot ECM in a limited fashion to take advantage of this offering.

Resolution, Compression & Image File Formats

Resolution

Image resolution is defined as the number of pixels per unit length, usually pixels per inch (PPI). Generally, the higher the number of PPI, the greater the legibility of the reproduced image. Images scanned at higher PPI rates use more storage space and may require longer scanning and retrieval times.

The selection of image resolution involves a trade-off between image clarity, storage capacity and speed. To determine proper resolution, you must first establish the desired quality of your images and the storage capacity of your information technology infrastructure. You will also need to consider the desired speed of delivery of images, especially if they will be accessed over the Internet. You may want to scan high-resolution masters of your images and create lower-resolution copies for Web delivery.

See Minimal Standards for Archival-Quality Images below and the table in Appendix A for resolution best practices.

Compression

Data compression is the process of encoding information using fewer bits. Image compression is the application of data compression to digital images. Compression methods may be used to reduce the amount of data needed to store or transmit images. There are two types of compression: lossless and lossy. Under lossless compression data is stored more efficiently but no data is lost. Under lossy compression, some data is permanently discarded. Lossy compression attempts to eliminate redundant or unnecessary information. Depending upon the degree of compression and other factors, the difference may be unnoticeable to the human eye.

For example, it is possible for a Joint Photographic Experts Group (JPEG) file (a lossy compression) and a Tagged Image File Format (TIFF) file (lossless) to appear exactly the same, although the JPEG file is missing data which makes it significantly smaller. These file formats and others are discussed below.

For a detailed explanation of compression, see the National Archives of the United Kingdom's Digital Preservation Guidance Note 5: Image Compression.

For compression standards, see ANSI/AIIM TR33-1998: Selecting an Appropriate Image Compression Method to Match User Requirements.

Image File Format

A file format is a particular way to encode information for storage in a computer file. A file format is usually described as either proprietary or nonproprietary. Proprietary file formats are controlled and supported by just one software developer, while nonproprietary formats are supported by more than one developer and can be accessed with different software systems.

Common types of digital image file formats include the following:

- Tagged Image File Format (TIFF) files are widely used in many different software programs and across all major operating systems. TIFF files are also free of patent and license issues. The file format utilizes lossless compression and is commonly used for master copies. TIFF graphics can be any resolution and can be black and white, grayscale, or color. The TIFF format has support for correct color space through International Color Consortium (ICC) profile documentation, which is an important consideration for art documentation.
- Joint Photographic Experts Group 2000 (JPEG 2000) files can utilize lossless or lossy compression for color and grayscale images. The compression method used is superior to the standard JPEG file listed below. Many institutions use this format to store master files, especially larger files, as JPEG 2000 files require less space than TIFF files. JPEG 2000 is also replacing the use of MrSID files. JPEG 2000 files can contain the same ICC color profile documentation as TIFF files.
- Joint Photographic Experts Group (JPEG) files use a lossy compression technique for color and grayscale images. Depending on the degree of compression, the loss of detail may or may not be visible to the human eye.

- Graphics Interchange Format (GIF) files support color and grayscale. Limited to 256 colors, GIFs are more effective for images such as logos and graphics than for color photos or art. GIF is a lossless compression. It should be noted that although the GIF format is widely used, it is technically proprietary.
- Portable Network Graphics (PNG) file format was designed to replace GIF files. PNG files can be ten to thirty percent more compressed than GIFs. Also utilizing lossless compression, PNG is completely patent and license free and is of higher quality than GIF.
- Bitmap (BMP) files are relatively low quality and used most often in word processing applications. BMP format creates a lossless compression.
- Portable Document Format (PDF) files are used to capture, distribute, and store formatted, page-oriented documents containing fonts, graphics, and images uncompressed. This format retains the "look" of the original document and can also include metadata. The format enjoys wide support in public and private institutions and Adobe (the creator of the PDF format) has licensed the patents for software that produces, consumes and interprets PDF files royalty-free to promote its use. The PDF version preferable for long-term storage is PDF/A, an open standard backed by the International Standards Organization (ISO).

Public and private institutions are currently using a mix of file formats for long-term storage. When color fidelity and fine detail are important, uncompressed TIFF files are the best option for long-term access and maintenance. When storage space is limited, consider using the JPEG 2000 format with lossless or lossy compression methods to significantly reduce the size of files that are oversized and/or do not require fine detail. In addition, institutions are beginning to use the PDF/A format for long-term storage of formatted, page-oriented documents that can contain fonts, images, and graphics. ISO 19005-1:2005 is the current standard describing PDF/A files. Standard JPEG files, PNG files and GIF files are appropriate formats for access and thumbnail images.

Minimal Standards for Archival-Quality Images

The table in Appendix A outlines the minimal standards required to achieve archival-quality digital images from scanned documents, artifacts, photographs and other media. Different original media types will require different conversion techniques as well as different file storage formats. Adhering to these minimal standards will ensure that the digital master files will record all of the significant visual features in the original item. Capture resolutions in the table are based upon the assumption that a scanning resolution of 300 PPI will be sufficient to meet this requirement for most originals in most collections, not including negatives and transparencies or slides.

Transmissive formats, such as negatives and slides, have a resolution standard of 3000 to 6000 pixels on the longest side, which yields an image of 300 PPI to 600 PPI when enlarged to 8” x 10” (3000 to 6000 pixels on the 10” long side).

The reflective formats, such as photographic prints and illustrations, are based on 8” x 10” originals scanned at 300 PPI. The 35mm film format has a resolution standard of 3000 to 6000 pixels in the longest dimension, as this is about as much data as most 35mm films can capture. Scanning the 35mm format, which is 1.5” on the longest side, at 3000 PPI will result in a file that can print an 8” x 10” item at 300 PPI.

Using film intermediaries as the main source for imaging is not recommended, as there is greater potential for loss with each derivation from the original document.

Indexing

Complete and accurate indexes are an essential component of an imaging system. Proper indexing using a standard taxonomy provides for efficient retrieval, ease of use, and up-to-date information about digital images stored in the system. Appropriate indexing methods include databases, spreadsheets, full-text optical character recognition (OCR) systems, document profiles and file naming conventions. When determining the index attributes to associate with the scanned documents, the following potential needs should be considered:

- Ready access to a record or logically associated set of records (e.g., the complete contents of a case file)
- Identification of documents by type (e.g., application, contract, invoice, etc.) and possibly by group (e.g., medical documents, correspondence, etc.) to allow for direct access to a document or logically related set of documents
- For case files (e.g., contract files, human resource files, etc.), a common identifier that will logically group all documents related to the case file together
- Where the imaging system is to be linked to a legacy system, an attribute that links to an attribute in the legacy database
- A date that establishes the start date for the retention period—for event-based retention periods, consider leaving a “blank” date field that can be filled in either programmatically or manually at the time the event occurs
- An attribute that may support establishing access security and restrictions at the document level, such as an individual record or group of medical records within a case file

- An attribute that may help facilitate automated routing of the documents to a particular workflow or process step

Backup, refreshing and data migration must ensure the preservation of all indexing associated with records in the imaging system, as well as the continued ability to identify, retrieve and reproduce all relevant documents.

Quality Control & Quality Assurance

Agencies and localities should assemble a sample set of source documents, or records equivalent in characteristics to the source documents, for the purposes of evaluating scanner results against defined quality criteria prior to production. Quality control criteria should be based upon the results of the pre-production quality sample. Quality criteria may include:

- Overall legibility
- Smallest detail legibility capture
- Completeness of detail
- Dimensional accuracy compared with the original
- Scanner-generated speckle
- Completeness of overall image area
- Density of solid black areas
- Color fidelity
- Image skew
- Image rotation
- Image cropping
- Index data accuracy
- Image and index format compliance

In addition, agencies and localities should adopt written quality assurance procedures for inspection of produced digital images. Quality assurance **must be** conducted before the original documents are destroyed.

Keep in mind that there is a significant difference between the quality control steps designed to detect and correct errors during the capture process and quality assurance that is designed to verify the validity and accuracy of the overall delivered product. While the capture process should provide quality control prior to product delivery, the end user must also perform his or her own quality assurance in order to verify that the delivered work product is acceptable.

Storage Media

Digital image file formats may require a great deal of physical storage, especially full-color files intended for archival storage purposes. Storage systems should be large enough to accommodate future growth and should also provide an appropriate level of certainty for the recovery and security of the images and related index attributes. In addition, it is important to develop backup procedures and policies regardless of the chosen storage media (See also: Section 3: System Implementation).

There exist three main types of digital media—magnetic, optical and solid state.

Magnetic media:

- Magnetic Disks include the hard disk found in your computer that stores the programs and files you work with daily. Also included are removable hard disks, external floppy disks, zip disks and removable cartridges. Magnetic disks provide random access.
- Magnetic Tapes come in reel-to-reel as well as cartridge format (encased in a housing for ease of use). The two main advantages of magnetic tapes are their relatively low cost and their large storage capacities (up to several gigabytes). Magnetic tapes provide sequential access to stored information, which is slower than the random access of magnetic disks. Magnetic tapes are a common choice for long-term storage or the transport of large volumes of information.
    - Digital Linear Tapes (DLT) come in a cartridge format a little larger than a credit card. Data is compressed using a special algorithm. DLT provides sequential access at high speeds.
    - Linear Tape-Open (LTO) is an open standard magnetic tape system. Similar to DLT in capacity and speed, LTO's standard format allows interoperability between tapes and tape drives made by different manufacturers.

Optical media:

- Compact Discs (CD) come in a variety of formats. These formats include CD-ROMs that are read-only, CD-Rs that can be written to once and are then read-only, and CD-RWs that can be written to in multiple sessions.
- Write-Once, Read-Many (WORM) Disks require a specific WORM disk drive to enable the user to write to or read the disk. WORM disks function the same way as CD-Rs.

- Digital Versatile Discs (DVD) are optical disks with more storage capacity than CD-ROMs. Common types of DVDs include DVD-ROM, DVD-RAM, DVD-R, DVD+R, DVD-RW, and DVD+RW.

Solid state devices:

- Flash-based Solid State Drives use flash memory rather than conventional spinning platters to store data.
- CompactFlash, SmartMedia, or Memory Sticks are most often found in digital cameras.
- PCMCIA Type I and Type II Memory Cards are used as solid-state disks in laptop computers.

Where data longevity or records integrity is a primary concern, non-rewritable media should be used. In addition, due to the limited life expectancy of digital media, no digital storage medium is adequate for the long-term or archival preservation of records. Assume that files need to be migrated to new storage media on a regular basis.

The United States National Archives and Records Administration makes the following storage recommendations in its *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*:

> We recommend that production master image files be stored on hard drive systems with a level of data redundancy, such as RAID drives, rather than on optical media, such as CD-R. An additional set of images with metadata stored on an open standard tape format (such as LTO) is recommended (CD-R as backup is a less desirable option), and a backup copy should be stored offsite. Regular backups of the images onto tape from the RAID drives is also recommended. A checksum should be generated and should be stored with the image files.
>
> Currently, we use CD-ROMs for distribution of images to external sources, not as a long-term storage medium. However, if images are stored on CD-ROMs, we recommend using high quality or "archival" quality CD-Rs (such as Mitsui Gold Archive CD-Rs). The term "archival" indicates the materials used to manufacture the CD-R (usually the dye layer where the data is recording, a protective gold layer to prevent pollutants from attacking the dye, or a physically durable top-coat to protect the surface of the disk) are reasonably stable and have good durability, but this will not guarantee the longevity of the media itself. All disks need to be stored and handled properly. We have found files stored on brand name CD-Rs that we have not been able to open less than a year after they have been written to the media. We recommend not using inexpensive or non-brand name CD-Rs, because generally they will be less stable, less durable, and more prone to recording problems. Two (or more) copies should be made; one copy should not be handled and should be

stored offsite. Most importantly, a procedure for migration of the files off of the CD-ROMs should be in place. In addition, all copies of the CD-ROMs should be periodically checked using a metric such as a CRC (cyclic redundancy checksum) for data integrity. For large-scale projects or for projects that create very large image files, the limited capacity of CD-R storage will be problematic. DVD-Rs may be considered for large projects, however, DVD formats are not as standardized as the lower-capacity CD-ROM formats, and compatibility and obsolescence in the near future is likely to be a problem.[2]

For additional information regarding digital media, see Electronic Records Guidelines.

Migration

Agencies and localities must ensure their long-term and permanent records are continually accessible. Therefore, imaging systems must enable the ongoing process of migration from older to newer hardware and software platforms. Current strategies for migrating records include:

- Upgrading equipment and software as technology evolves
- Recopying media based upon projected longevity and/or periodic verification of the records
- Transferring the data from an obsolete medium to a newly-emerging technology, in some cases bypassing the intermediate generation that is mature but at risk of becoming obsolete

[2] Puglia , Steven, Jeffrey Reed, and Erin Rhodes. *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*. National Archives and Records Administration, June 2004, p. 60.

DIGITAL IMAGING GUIDELINES
SECTION 3: SYSTEM IMPLEMENTATION

An imaging system incorporates a myriad of tasks, the successful management of which can save time and money. While a vendor may be contracted for the project, agency or locality staff will need to manage an assortment of activities, including the:

- Selection of materials to be digitized
- Preparation of materials, including sorting files, removing staples and paperclips, weeding out unnecessary materials, and conserving any deteriorating documents
- Creation of standardized indexing attributes (or metadata)
- Quality control of source materials and digital images
- Staff training on new hardware and software
- Advertising, promotion and user evaluation
- Long-term maintenance of images

The following recommendations should be considered during system implementation:

1. **Establish Operating Procedures and Provide Documentation**

   Establish operational practices and provide technical and administrative documentation to ensure the future usability of the system, continued access to long-term records and a sound foundation for assuring the system's legal integrity. Documentation of operating procedures should include a description of methods for scanning, entering data, revising, updating, and expunging records, indexing techniques, and backup procedures, as well as hardware and software operating manuals. Documentation should also include procedures for testing the readability of records, security safeguards to prevent tampering with and unauthorized access to protected information, and the disposition of original records.

2. **Institute Procedures to Ensure Quality Integrity of Images**

   Include visual inspection in your operational procedures to verify the completeness and accuracy of the scanning process. In addition, the procedures should include inspection of images to confirm that storage media and other system components function properly. Audits should be performed regularly, particularly following installation of new software or

hardware upgrades. Each audit should produce a written report detailing the date of the audit, the name of the auditor and the findings.

3. Apply Retention and Disposition Schedules

Public officials are responsible by law for ensuring that their records—paper and electronic—are preserved, maintained and accessible throughout the record's lifecycle. The lifecycle of a record is determined by its records retention and disposition schedule, of which there are two kinds; "general schedules" apply to records commonly found in most agencies or localities and "specific schedules" apply to records that are unique to an agency or locality. State Agency General Schedules and Locality General Schedules are available on the Library of Virginia's (LVA) Web site.

The LVA's Archival and Records Management Services Division issues retention and disposition schedules that specify how long various records must be kept. Each state agency and locality is required to designate at least one records officer to serve as a liaison to the LVA for the purposes of implementing a records management program. Because many records management issues must be considered when planning and implementing an imaging system, records officers must be involved throughout the process.

Images should be stored in such a way that they can be identified and disposed of in the appropriate manner at the appropriate time. It is particularly important to distinguish records containing identifying (or privacy-protected) information, defined by the Virginia Public Records Act of the *Code of Virginia* (42.1-86.1) to include: social security numbers, driver's license numbers, bank account numbers, credit or debit card numbers, personal identification numbers (PIN), electronic identification codes, automated or electronic signatures and passwords. Records containing such information must be destroyed within six months of the expiration of the record's retention period. Furthermore, electronic records containing identifying information must be electronically shredded, as opposed to deleted.

The application of retention and disposition schedules can be facilitated by an integrated electronic document management (EDM) system and electronic records management (ERM) system, also known as an Enterprise Content Management (ECM) system. For more information about EDM, ERM and ECM, see Section 2: Technical Specifications and Selection.

4. Provide Adequate Environmental Conditions

Digital media is susceptible to deterioration when storage conditions are inadequate. The Association of Records Managers and Administrators (ARMA) suggests a temperature of 65 to 70 degrees Fahrenheit and a relative humidity of between 20 and 35 percent for digital media storage. Media should never be in direct sunlight or near heat sources, and should be protected from dust, debris, fingerprints and high-static environments.

5. Establish Backup Procedures

The purpose of backup is to provide replacement of data or images lost due to system or user error, or in the event of a disaster. This is especially vital in terms of imaging, as the original paper-based information may no longer exist. Backup and restoration of data requires the following to be strictly enforced:

- Documented procedures for backups
- Regular audits of the procedure to determine validity and completeness of data and images to be restored

6. Establish Continuity of Operations Plan

A records disaster is a sudden and unexpected event that results in the loss of records or information essential to an organization's continued operation. Disasters include fire, flood and tornado. Less obvious, but equally disastrous, events include human error, vandalism, unauthorized access, loss, theft, equipment failure, leaking pipes, insects, rodents, mold and terrorism. Continuity of Operations Planning (COOP), or disaster preparedness planning, is critical to the Commonwealth's ability to deliver valuable services to its citizens during and immediately after a disaster.

COOP for imaging systems should include provisions for:

- Off-site storage of index data, digital images and system documentation
- Adequate security for backup media
- A recovery plan
- A plan for routine system audits

In addition, COOP should include a vital records plan to safeguard those records that are essential to the continued functioning or reconstitution of a government entity.

Plans should be tested periodically to validate effectiveness and to identify areas that could be strengthened. They should be tested at least annually, with portions of the plan subject to more frequent testing if appropriate.

DIGITAL IMAGING GUIDELINES

SECTION 4: ARCHIVING AND LONG-TERM MAINTENANCE

Public officials are responsible by law for ensuring that public records are accessible throughout their lifecycles. According to the Virginia Public Records Act of the *Code of Virginia (*42.1-85), this responsibility includes the conversion and migration of electronic records "as often as necessary so that information is not lost due to hardware, software, or media obsolescence or deterioration."

Hardware, software and file formats can be operational for ten years or more, but technology is often superseded within two or three years. If a system stores records with retention periods exceeding the lifespan of the hardware or software, it is essential that the system administrator plan for data sustainability. A digital sustainability strategy documents how an organization will transfer long-term or permanent records from one generation of hardware and software to another while maintaining system functionality and data integrity. The strategy should include current system documentation and should be updated as technology changes.

Please consult with your records analyst at the Library of Virginia (LVA) prior to reformatting permanent records. Original records that have a permanent retention **must be** offered to the LVA prior to being destroyed.

If digital images need to be maintained for long periods of time, i.e., more than ten years, it will be necessary to address the following issues to ensure accessibility:

1. **Scanning Specifications and File Formats**

   For images that need to be retained ten years or more, it is recommended that originals be scanned at a minimum of 300 PPI and saved as uncompressed Tagged Image File Format (TIFF) files. Best practice indicates that it is preferable to use the most recent version of the file format. The TIFF image will serve as a "master image" or "archival copy" that is similar to a microfilm master negative. The master image should capture as much information as possible from the original in order to serve as a long-term, high-quality digital version. The TIFF images should be stored uncompressed for maximum accessibility, now and in the future. If it is absolutely necessary to compress the master image files, follow current industry standards provided by ANSI/AIIM TR33-1998: Selecting an Appropriate Image Compression Method to Match User Requirements. The TIFF images should be stored in a secure and stable environment, preferably off-line.

Where storage space is limited, the Joint Photographic Experts Group 2000 (JPEG 2000) file format with lossless or lossy compression is appropriate to use as a master image when images are oversized and/or do not require fine detail. As with TIFF master images, the original JPEG 2000 masters should be stored in a secure and stable environment.

The archival Portable Document File (PDF/A) format is becoming acceptable to use as a master image for storing formatted, page-oriented documents that can contain fonts, graphics, and images. As with TIFF and JPEG 2000 master images, the original PDF/A masters should be stored in a secure and stable environment.

Derivative images for use on the Web or within the application itself should be made from the master image. Master images can be bitonal, grayscale or color. Quality control applied to master images should be intensive. This is especially true if an agency or locality is retaining only a digital image of a record with permanent retention.

2. Metadata

Metadata consists of a standardized structured format and controlled vocabulary that allow for the precise description of record content, location, and value. Metadata often includes, but is not limited to, attributes like file type, file name, creator name, date of creation and use restrictions. Metadata capture, whether automatic or manual, is a process built into the information system.

See Section 2: Technical Specifications and Selection for additional information regarding metadata and indexing.

Also refer to the following metadata standards:

- Dublin Core Metadata Initiative's Metadata Terms
- Department of Defense 5015.2-STD: Design Criteria Standard for Electronic Records Management Software Applications
- International Organization for Standardization (ISO) 15489: Information and Documentation – Records Management

3. System Information and Maintenance

On a system level, documentation consists of information about planning, development, specifications, implementation, modification and maintenance of system components (hardware, software, networks, etc.). System documentation includes policies, procedures, data models, user manuals and program codes.

Please refer to Section 9 of the Minnesota State Archives' Trustworthy Information Systems Handbook for additional guidance on system documentation.

4. Migration and Conversion

The most common approach to preserving electronic records involves a combination of two techniques: migration and conversion. Migration is the process of moving files to new media (also known as "media refreshing") or computer platforms in order to maintain their value. Conversion entails changing files from one format to another and may involve moving from a proprietary format to a nonproprietary one. To avoid losing data in the process of migration or conversion, you should perform initial tests and analyses to determine exactly what changes will occur and whether they are acceptable. With both migration and conversion, special attention must be paid to maintaining the accessibility of associated metadata. When properly planned and executed, the migration and conversion approach probably represents the easiest and most cost-effective preservation method available today.

DIGITAL IMAGING GUIDELINES

APPENDIX A: MINIMAL STANDARDS FOR ARCHIVAL-QUALITY IMAGES TABLE

Media Type	Imaging Method	Spatial Master Resolution	Master File Format	Access Image File Format and Resolution	Thumbnail Image File Format and Resolution
Photographs and Small Pictorial and Graphic Materials	Small flatbed scanner or digital camera	3000 to 6000 pixels across long dimension	24-bit RGB color or 8-bit grayscale TIFF or JPEG 2000	24-bit RGB color or 8-bit grayscale JPEG at 150 DPI	24-bit RGB color JPEG, 8-bit indexed color GIF, or 8-bit grayscale at 72 DPI
Large Graphic Materials (e.g., maps, posters, broadsides, unbound manuscripts, architectural drawings)	Large flatbed scanner or scanning back in copy stand or medium to large format camera	3000 to 12000 pixels across long dimension	24-bit RGB color TIFF or JPEG 2000	24-bit RGB color JPEG at 150 DPI or MrSID format	24-bit RGB color JPEG or 8-bit indexed color GIF at 72 DPI
Bound Material (e.g., text materials reproduced as images, atlases, folios, bound newspapers)	Digital camera or scanning back in copy stand	3000 to 12000 pixels across long dimension	24-bit RGB color or 8-bit grayscale TIFF or JPEG 2000	24-bit RGB color or 8-bit grayscale JPEG at 150 DPI	24-bit RGB color JPEG, 8-bit indexed color GIF, or 8-bit grayscale at 72 DPI
Text Materials (e.g., reproduced as searchable text)	Flatbed or document feeder scanner	300 to 600 PPI	Determined by OCR software	N/A	N/A
Transparencies and Negatives	Slide scanner for small formats; drum or transparency scanner for medium to large formats	3000 to 6000 pixels across long dimension	24-bit RGB color or 8-bit grayscale TIFF or JPEG 2000	24-bit RGB color or 8-bit grayscale JPEG at 150 DPI	24-bit RGB color JPEG or 8-bit indexed color GIF at 72 DPI
Paintings and 3-Dimensional Objects (e.g., sculpture)	Digital camera or scanning back in medium or large format camera	3000 to 12000 pixels across long dimension	24-bit RGB color TIFF or JPEG 2000	24-bit RGB color JPEG at 150 DPI or MrSID format	24-bit RGB color JPEG or 8-bit indexed color GIF at 72 DPI

Media Type	Imaging Method	Spatial Master Resolution	Master File Format	Access Image File Format and Resolution	Thumbnail Image File Format and Resolution
Microfilm (35mm, 16mm, and microfiche)	Microfilm scanner	200 – 400 DPI (capable of filling letter-size sheet when printed)	Bitonal or grayscale TIFF or JPEG 2000	Bitonal or grayscale JPEG	Bitonal or grayscale JPEG or TIFF

DIGITAL IMAGING GUIDELINES

APPENDIX B: GLOSSARY

Administrative value: Records shall be deemed of administrative value if they have continuing utility in the operation of an agency or locality.

Archival quality: A quality of reproduction consistent with established standards specified by state and national agencies and organizations responsible for establishing such standards, such as the Association for Information and Image Management, the American National Standards Institute or the National Institute of Standards and Technology.

Archival record: Material created or received in the conduct of affairs that is preserved because of the enduring historical value or as evidence of the roles and responsibilities of the creator.

Backup: A copy of an electronic record maintained to protect information loss.

Barcode: A coding system composed of vertical lines set in patterns that, when read by an optical reader, can be converted into electronic text.

Barcode recognition (BCR): The recognition of barcodes by a computer and the conversion of the barcodes into electronic text to create an index for images.

Bitonal: One bit per pixel representing black and white. Bitonal scanning is best suited to high-contrast documents such as printed text.

Bits: Short for binary digits, the smallest unit of information in a binary system. Each bit is assigned a 1 (high current) or a 0 (low current), where eight bits make up a byte.

Color: Multiple bits per pixel representing color. Color scanning is suited to documents with color information.

Continuity of operations plan (COOP): An internal effort within an organization to ensure the ability to continue essential business functions across a wide range of potential emergencies.

Conversion: The act of moving electronic records to a different format, especially data from an obsolete format to a current format.

Compression: A computer process using algorithms that reduces the size of electronic files so they occupy less digital storage space.

Data compression: Any of several techniques that reduce the number of bits required to represent information in data transmission or storage.

Derivative image: An image that is used in place of a master image for general Web access. Includes “thumbnail” and “reference” or “service” images. Images created for this purpose commonly have smaller file sizes and are in a Web-viewable format such as JPEG or GIF.

Digital image: See image.

Digital imaging: See imaging.

Disposition: Action to be taken on a records series at a specific time. May entail destruction, reformatting, transfer or permanent retention.

Dots per inch (DPI): A measurement of the scanning resolution of an image or the quality of an output device. It expresses the number of dots per inch, both horizontally and vertically, that a printer can print or that a monitor can display.

Enterprise content management (ECM) system: Technologies used to capture, manage, store, preserve and deliver content and documents related to organizational processes. Electronic Document Management (EDM) and Electronic Records Management (ERM) are components of ECM.

Electronic document management (EDM) system: Software that controls the capture, indexing, processing, storing, transferring and use of electronic documents to facilitate workflow. Manages documents as individual units, as opposed to preserving relationships to larger groups of documents that provide evidence of the same particular organizational function.

Electronic records management (ERM) system: Software that enables the capture and management of electronic documents as records. Typical ERM functions include declaration, capture, organization, security, retrieval, preservation, audit/oversight and disposition.

Electronic shredding: The process of overwriting data instead of merely deleting it. Involves overwriting the file's data clusters, renaming the file with a randomly generated name, truncating the file to 0 bytes in length and deleting the renamed and truncated file.

File format: A specification for organizing data. Digital images and their associated metadata may be presented in a number of formats depending on compression schemes, intended use, or interoperability requirements. Some image formats are broadly decipherable, while others may only be accessible to certain application programs.

Film intermediary: A surrogate or derivative of an original document that has been reformatted to film. Film intermediaries of originals include 35mm slides, 4" x 5" transparencies, microfilm, and single-frame microfiche.

Fiscal value: Records shall be deemed of fiscal value if they are needed to document and verify financial authorizations, obligations and transactions.

Grayscale: Multiple bits per pixel representing shades of gray. Grayscale is suited to continuous tone documents, such as black and white photographs.

Historical value: Records shall be deemed of historical value if they contain unique information, regardless of age, that provides understanding of some aspect of the government and promotes the development of an informed and enlightened citizenry.

Identifying (or privacy-protected) information: According to Section 18.2-186.3 of the *Code of Virginia*, identifying information includes social security numbers, driver's license numbers, bank account numbers; credit or debit card numbers, personal identification numbers (PIN), electronic identification codes, automated or electronic signatures and passwords.

Image: A graphic representation of an object. More specifically, a raster-based, two-dimensional, rectangular array of static data elements called pixels, intended for display on a computer monitor or for transformation into another format, such as a printed page.

Image compression: The application of data compression on digital images.

Imaging: The process of electronically capturing the visual appearance of (usually) paper documents, also called "scanning" or "digitizing."

Index: Descriptive data associated with an image for retrieving that specific image from storage.

Legacy system: An existing computer system that must be accommodated when building new systems.

Legal value: Records shall be deemed of legal value if they document actions taken in the protection and proving of legal or civil rights and obligations of individuals and agencies.

Lifecycle: The creation, use, maintenance and disposition of a public record.

Lossless compression: Reduction in file size without loss of information, achieved by storing data more efficiently.

Lossy compression: Reduction in file size that involves permanent loss of information. Algorithms selectively discard data in order to attain a greater size diminishment than is possible with lossless compression.

Magnetic media: Tape or disk coated with a magnetic surface used for storing electronic data.

Master image: A faithful digital reproduction of a document optimized for longevity and for production of a range of delivery versions.

Metadata: Data describing the context, content and structure of records and their management through time.

Micrographics: The techniques associated with the production and handling of microfilm, microfiche and related storage technologies based on retaining a photographic representation on film.

Migration: The process of moving records from one hardware and/or software platform to another.

Nonproprietary: A format that is not owned by a private individual or corporation under a trademark or patent. It is in the public domain and is easily portable between various hardware and software systems.

Optical character recognition (OCR): The recognition of printed or written text characters by a computer. Involves photo scanning of the text character-by-character, analysis of the scanned image and translation of the character image into character codes, such as ASCII, commonly used in data processing.

Optical media: A data storage medium that utilizes laser technology to read information.

Pixel: Short for picture elements, which make up an image, similar to grains in a photograph or dots in a halftone. Each pixel can represent a number of different shades or colors, depending on how much storage space is allocated for it. Pixel size, frequency and color determine the accuracy with which photographic images can be represented.

Proprietary: A format that is owned by a company or a private owner. Some proprietary formats are published and protected by intellectual property rights or copyright. Other proprietary formats are not published.

Public record: Recorded information that documents a transaction or activity by or with any public officer, agency, or employee of an agency. Regardless of physical form or characteristic, the recorded information is a public record if it is produced, collected, received, or retained in pursuance of law or in connection with the transaction of public business. The medium upon which such information is recorded has no bearing on the determination of whether the recording is a public record.

Quality assurance: The process by which the total product is examined to ensure that the quality criteria initially established in the pre-production test has been met.

Quality control: Techniques to ensure accuracy and high quality through various stages of a process.

Record copy: An original, official, or master record that is distinct from a "working" or "convenience" copy, which is a duplicate used for reference purposes.

Records analysis: The examination and evaluation of systems and procedures related to the creation, processing, storage and disposition of records.

Records retention and disposition schedule: A Library of Virginia-approved timetable stating the required retention period and disposition action of a records series. The administrative, fiscal,

historical, and legal value of a public record shall be considered in appraising its appropriate retention schedule.

Refresh: The process of periodically moving records from one storage medium to another.

Resolution: The measure of the quality of a digital image, usually expressed in dots per inch (DPI).

Retention period: The length of time a record is kept.

Scanning: See imaging.

Solid state media: A data storage medium that uses solid-state memory with no moving parts.

Standards: Rules typically developed, adopted and promoted by large organizations that can advocate for their broad usage. Data standards enable the exchange of data while technology standards enable the delivery of data between systems.

Taxonomy: A set of controlled vocabulary terms, usually hierarchical.

Vital record: A record essential to the operation of an organization and/or resumption of operations following a disaster.

Workflow analysis: The examination and evaluation of the tasks, procedural steps, staff involved, required input and output information, and tools needed for each step in a business process.