

HISTORICAL & ARCHIVAL RECORDS CARE GRANTS: Digitization Guidelines for Textual and Graphical Records

Digitization projects should result in the files described below and meet the parameters defined herein. With proper software it is not necessary to subject source materials to multiple scans. Derivative files should be generated from the Master Image File. Resulting files should include:

A. For projects where the original materials *will* be retained:

1. Master Image File: TIFF (graphical) or PDF/A (textual)

Proper digital imaging methodology involves scanning a high-quality master or archival image from which versions in smaller sizes or alternative formats are derived for a variety of uses. A digital master must be available and rich enough to accommodate future needs and applications. The master image should be the highest quality you can afford; it should not be edited nor processed for any specific output; and it should be uncompressed. Intensive quality control should be applied in creating master image files. Any errors made while creating the master file will necessitate going back to the scanner or camera to capture another image.

- Textual Material: PDF/A or uncompressed TIFF.
- Graphical Material such as photographs and drawings: uncompressed TIFF.

2. Access File: JPEG, TIFF or PDF/A

This Access File, derived from the Master Image File, can be used for general reference purposes, distributed, and altered as required. JPEG, TIFF or PDF/A format ensure that copies of scanned materials can be uploaded to PA Power Library.

B. For government projects where the original materials *will not* be retained, a Security Preservation File and Backup File must also be produced and maintained as outlined in the County Records Manual:

1. Security Preservation File: PDF/A

The Security Preservation file is the copy of an electronic record that is stored in a secure manner so that it cannot be accessed except by a limited number of authorized users and only when no other copy of the document will suffice. This file must be retained in accordance with PHMC policy in a live online environment and must not be

retained solely on any kind of removable media. For instances where a government project has scanned textual only materials to a PDF/A Master Image File, this file may also be used as the Security Preservation File. Projects involving graphical materials (photographs, drawings, etc.) should consult with the State Archives about appropriate preservation formats.

2. Backup File: PDF/A

To be used only in the event of a disaster that disrupts or destroys the remaining files should be retained at a site that is geographically separated from the original Security Preservation Files.

Scanning Parameters

Minimum Scanning Resolution and Bit Depth

Textual Documents, Manuscripts, and Bound Volumes, Printed

Resolution	300 ppi
Bit Depth	8 or 16 bit grayscale

Textual Documents, Manuscripts, and Bound Volumes, Hand Written

Resolution	300 ppi
Bit Depth	24 bit RGB

Mixed (text and image) Documents, Manuscripts and Bound Volumes

Resolution	300 to 600 ppi
Bit Depth	24 bit RGB

Oversize Items: Maps, Posters, and Other Materials

Resolution	300 ppi
Bit Depth	8 or 16 bit grayscale; 24 bit RGB

Prints and Photographs

Resolution	430 ppi (11x14); 600 ppi (8x10); 800 ppi (3x5)
Bit Depth	8 or 16 bit grayscale (for black and white images); 24 bit RGB

Photographic Negatives: 35mm to 4"x5"

Resolution	3000 ppi
Bit Depth	16 bit grayscale or 24 bit RGB as appropriate

Terms

Image Editing Software: Once you have captured an image from your scanner and have saved an unaltered master file to your storage media, you will need to manipulate surrogate copies of that image using image editing software. Staff will need professional image editing software.

Image Resolution: Resolution determines the quality of an image. It is described either by pixel dimensions (height and width) for on-screen use or by physical size and PPI. Increased PPI take more frequent samples of the original and contain a more accurate representation of the original. Since higher resolutions are capturing more information, file sizes also increase. There is no one “perfect” resolution to scan all collection materials. Resolution should be adjusted based on the size, quality, condition and uses of the digital object. The combination of PPI and size of the original object determine the resolution needed to accurately capture as much information about the original object as is available. There is a point at which adding more pixels per inch no longer adds content, because the original source object has a finite amount of information available based on the way that it was produced.¹

JPEG (Joint Photographic Experts Group): A lossy compression technique for color and grayscale images. Depending upon the degree of compression, the loss of detail may or may not be visible to the human eye. Files in JPEG format end with a .jpg extension. JPEG image files require less storage and are quicker to download.

PDF/A (Portable Document Format / Archives): A variant of PDF that is specifically aimed at long-term preservation, its specifications are published in the standard ISO 19005-1:2005. It sacrifices certain functions, such as the ability to have external hyperlinks or embed audio or video, for the sake of greater reliability. The most notable difference between PDF and PDF/A is the latter’s ability to embed all necessary fonts within the file itself. This makes the file totally self-extracting, without any need to access external font information to properly present the formatting of the document. PDF/A also embeds descriptive metadata within the file itself, making it self-describing. These two factors make PDF/A the preferred format for long-term preservation of textual electronic records, both born-digital and digitized. Files can be converted to PDF/A by a number of different software tools and plug-ins to existing word-processor software. PDF/A has been chosen as the Security Preservation File format by the County Records Committee.

PPI/DPI: Pixels per inch (PPI) refers to the number of pixels captured in a given inch and is used when discussing scanning resolution and on-screen display. When referring to digital capture, PPI is the preferred term, as it more accurately describes the digital image. Digitization guidelines and hardware manufacturers’ documentation frequently use the measurement of dots per inch (DPI) when discussing optical resolutions for images and hardware. DPI more accurately refers to output devices, how many dots of ink per inch a printer puts on the paper or onscreen monitor display.²

RGB: RGB refers to multiple bits per pixel representing color. Color capture is suited to items with continuous tone color information. Manuscripts, older printed matter and sheet music may be better served by capturing as continuous tone in grayscale or color to bring out the shade and condition of the paper and the marks inscribed on it. Projects interested in capturing the current condition of source materials should consider capturing in color.³ Be aware that the combination of an increased bit depth and resolution will result in large to extremely large image files. These files may be difficult and/or costly to manage. Because of storage concerns and time considerations, it may be necessary to reduce the recommended resolution and/or bit depth when creating master images.⁴

TIFF (Tagged Image File Format) TIFF was initially created in the 1980s in an effort to standardize file formats created by commercial scanners. The format has gone through a number of revisions since then, becoming an international standard for electronic images. The format is currently owned by Adobe Corporation, but the specifications are open and freely available. Unlike many image file formats, TIFF is uncompressed. This means that the files are larger than a compressed format (such as JPEG) but there is no loss of data. This ensures that the file can be reproduced over time at its full fidelity. TIFF files can contain “tags” that store descriptive metadata about the file. TIFF files may have a file extension of .tif (Windows) or .tiff. Tiffs may be converted into PDF/A.

1. BCR’s CDP Digital Imaging Best Practices, Version 2.0, June 2008, Page 10
2. BCR’s CDP Digital Imaging Best Practices, Version 2.0, June 2008, Page 8

3. BCR's CDP Digital Imaging Best Practices, Version 2.0, June 2008, Page 8
4. BCR's CDP Digital Imaging Best Practices, Version 2.0, June 2008, Page 9

Minimum Metadata for POWER Library

Grantees are encouraged to submit digitized materials to the POWER Library or DPLA. At minimum, POWER Library requires that the collection be cataloged using Dublin Core Metadata. For more information, please review the Metadata Specifications on Pages 4 – 15 of the PA Power Library Guidelines at: https://www.powerlibrary.org/wp-content/uploads/2015/09/PAPhotos_guidelines.pdf.

Reference Documents

Atlanta Housing Authority, Scanning Documents and Photographs, A Proposed Guideline and Research Paper, 2012.

Bibliographical Center for Research (BCR) Colorado Digitization Program (CDP) Digital Imaging Best Practices, Version 2.0, June 2008. http://mwdl.org/docs/digital-imaging-bp_2.0.pdf

County Records Manual Issued for the County Records Committee By the Pennsylvania Historical And Museum Commission, updated April, 2017. <http://www.phmc.pa.gov/Archives/Records-Management/Documents/RM-2002-County-Records-Manual-2017-Update.pdf>

Federal Agency Digitization Guideline Initiative (FADGI) Technical Guidelines for Digitizing Cultural Heritage Materials, September, 2016. http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

Library of Virginia, DIGITAL IMAGING GUIDELINES, September 2008. http://www.lva.virginia.gov/agencies/records/electronic/digital_imaging.pdf

National Archives and Records Administration Scanning Guidelines, June 2004. <https://www.archives.gov/files/preservation/technical/guidelines.pdf>