

# **Digitizing Vertical Files**

## **at National Gallery of Art Library**

### **Project Planning and Management Guidelines**

**July 2013**

**Authored by Kai Alexis Smith**

# Table of Contents

Introduction .....	4
Purpose of the Vertical Files .....	4
Why Digitize the Vertical Files?.....	4
Acronyms and Definitions.....	4
Approaches to File Creation .....	5
Proposed Workflow/Process .....	5
Similar Frameworks and Attempts .....	6
Prioritization .....	7
Internal Tracking System .....	7
Handling Materials Guidelines.....	8
Background .....	8
Folder Labeling and Curator/Researcher Buy In .....	9
Digitization Workflow Documentation .....	10
Figure 1. 5 Phase Project Management Workflow .....	11
Digitization General Principles.....	11
Digitization Workflow .....	12
Figure 2. Recommended Scanning Process for Approach 1 .....	12
Figure 3. Recommended Scanning Process for Approach 2 .....	13
General Digitization Requirements.....	14
Digital Asset Management System (DAM).....	14
File Formats and Bit Depth .....	14
Resolution and Sizing .....	14
Color Management .....	15
Technical Requirements .....	15
Software Needed .....	15
Calculating Server Space .....	15
Metadata .....	15
File Naming .....	16
Multiple Pages in a Document .....	16
PDF/As .....	17
Future Changes .....	17

Creating Preservation Masters TIFF.....	18
Cataloging Addition .....	18
Image Manipulation Guidelines.....	19
Watermarks .....	19
Orientation .....	19
Cropping .....	19
Blank Pages .....	19
Copyright .....	20
Access .....	20
Technology Literacy .....	20
References .....	21
Table 1: Excel Spreadsheet documenting workflow of files .....	23

## Introduction

This document defines several approaches to digitizing the National Gallery of Art (NGA) Library's vertical files (VF) and outlines a series of high-level activities for planning and management purposes. While the outline is specific to the vertical files at the National Gallery of Art, these strategies can be applied to other museums and institutions that would like to make their collections more accessible. Digitization is defined as the creation of digital objects from physical materials. Born Digital materials are not currently within the scope of this document.

## Purpose of the Vertical Files

A valuable component of the NGA Library's monographs and serials holdings, the VF serves the research needs of the Gallery's staff members, fellows, visiting scholars, researchers and interns. The collection of exhibition catalogues and ephemera are especially useful for researchers of modern and contemporary art and under-recognized artists.

## Why Digitize the Vertical Files?

Providing digitized versions of the library's holdings furthers its mission to support the curatorial and research needs of the gallery's staff. In addition, as per the Information Technology Department (IT) Strategic Plan 2020, digitizing the VF collection takes a step toward developing the Gallery's intranet, which makes documents available internally. Digitized content will be available through the Library catalog Mercury, allowing for an increased level of access to the collection, while preserving originals by reducing wear and tear. Currently, there are 12,000 vertical files. Only 4,400 folders are cataloged and findable in the catalog. This plan will increase access significantly.

While increased access to the digitized materials will benefit the collection, the originals will not be de-accessed, moved off site, or destroyed after being scanned.

## Acronyms and Definitions

Artist Files	Sometimes called Vertical Files and vice versa
DAM	Digital Asset Management System
Digitization	Process of scanning of physical materials
DOI	Digital Object Identifier
JPG	Joint Photographic Experts Group, this is a file format
IT	Information Technology Department
NGA	National Gallery of Art
PDF/A	Archival version of Portable Document Format

Preservation Masters (PM)  
Surrogates  
TIFF  
VF

Phrase referring to TIFFs  
Phrase referring to JPGs  
Tagged Image File Format  
Vertical File

## Approaches to File Creation

There are two ways to approach the digitization of the vertical files.

**Approach 1:** This involves the creation of Preservation Master files (TIFFs), surrogates (JPGs), and PDF/As. This approach will ensure archival versions of items in the files for future use.



**Approach 2:** Scan the items into Adobe Acrobat and create compressed PDF/As. This method skips creating archival files and simply creates PDF/As for direct access.

## Proposed Workflow/Process

This document will divide the process into 5 phases:

Phase 1	Project Plan
Phase 2	Pre Digitization
Phase 3	Digital Conversion
Phase 4	Post Digitization
Phase 5	Access

Some of the ongoing activities throughout the phases will include project management and quality management.

## Similar Frameworks and Attempts

Early attempts to increase access to similar collections include the Artist Files Revealed project and the Smithsonian American Art Museum List of Artist Files. The Artist Files Revealed online directory is a centralized repository for cultural institutions to promote their collections of artist files and for users to discover these resources more readily.<sup>1</sup> It reveals contact information and provides information on each institution's artist files collections. The Smithsonian American Art Museum List of Artist Files provides information on the artist file holdings across the Smithsonian Institutions. However, the delivery of publicly accessible virtual documents is currently impossible because of copyright restrictions.

State-side, the Guggenheim Museum has been successful with digitizing and making materials from artist files available on their website. They have taken precautions with a detailed rights and restricted use policy, the images cannot be downloaded, and the quality of the images is not high enough resolution for publication.

The proposed approaches in this report are similar to what The Information Center for Israeli Art at the Israel Museum has done. Since April 2010, the staff has been able to successfully digitize physical files that are only accessible via an in-house database. However, they have taken it a step further by partnering with Wikipedia. Wikipedia entries in both Hebrew and English on about 5,800 Israeli artists provide yet another access point to their artist files and promote their collection.

All efforts to provide access to valuable collections such as artist files are admirable and each project continues to build upon the other.

---

<sup>1</sup>(2010). Artist Files Revealed. Retrieved from <http://www.artistfilesrevealed.com/tiki/tiki-index.php>

## Prioritization

The NGA Library's VF includes a variety of ephemera including announcements, brochures, exhibition catalogs, flyers, guidebooks, invitations, paperback monographs, obituaries, pamphlets, periodical literature, promotional posters and press releases. However, the NGA Library will only scan announcements, press releases, and pamphlets in an effort not to duplicate efforts. Anything five pages or more will not be scanned, but will be cataloged to comply with the VF policies and guidelines.

Policies for prioritization efforts are divided into 3 tiers:

### Tier 1

- Identify frequently requested or checked out folders.
- Identify missing items returned.
- Identify any folders recovered from curator's offices.
- Identify new acquisitions. If the folder has not been scanned, then the whole folder should be scanned.

### Tier 2

- Identify folders of modern and contemporary artists in up and coming exhibits.

### Tier 3

- Identify folders that will demonstrate a need to one or more curators, staff, or fellows.

## Internal Tracking System

Quantitative data is essential to determining and identifying folders to be scanned.

Developing an internal tracking system of items checked out is recommended. This can be managed in a spreadsheet in excel, through Google Forms, or possibly through Voyager if this feature is available. If Voyager is used, barcoding at the folder level is highly recommended.

Tracking who the folders are checked out to, which folders were checked out, and how often each folder is checked out is important. To facilitate this process faster it is recommended the Vertical Files librarian enter this data whenever a file is checked out and returned.





## Folder Labeling and Curator/Researcher Buy In

After the scanning process has been completed add NGA designed labels to the folders indicating that the folder was scanned and is available in the catalog. Here are two examples of the NGA seal that could be made into a sticker:



In an effort to increase awareness of the project, a poster of any size should be created to let curators know about the NGA seal on the folders indicating that they are scanned and invite them to request folders for scanning. Language on the poster is suggested below:

If you see the NGA seal on a folder, that means the folder has been scanned and is available in Mercury.

The VF department has been digitizing folders. Help us out.

Do you have a folder in your office or do you have a folder in mind you would like scanned?

Make a request by asking [INSERT NAME].

Folder scans may take between [Insert time frame here]. Plan accordingly.

Thank you for your patience.

This approach will not only notify curators and researchers about the digitization efforts, but also involve them in the process. Be clear to indicate the average time it will take to scan the folder.

## Digitization Workflow Documentation

There are over 12,000 folders in the VF collection; however, the numbers of physical items inside each folder is not known. Developing a document to track the digitization of objects is vital to understanding method, where folders are in the digitization process, recording decisions, tracking problems, and helping to maintain consistency and quality control.

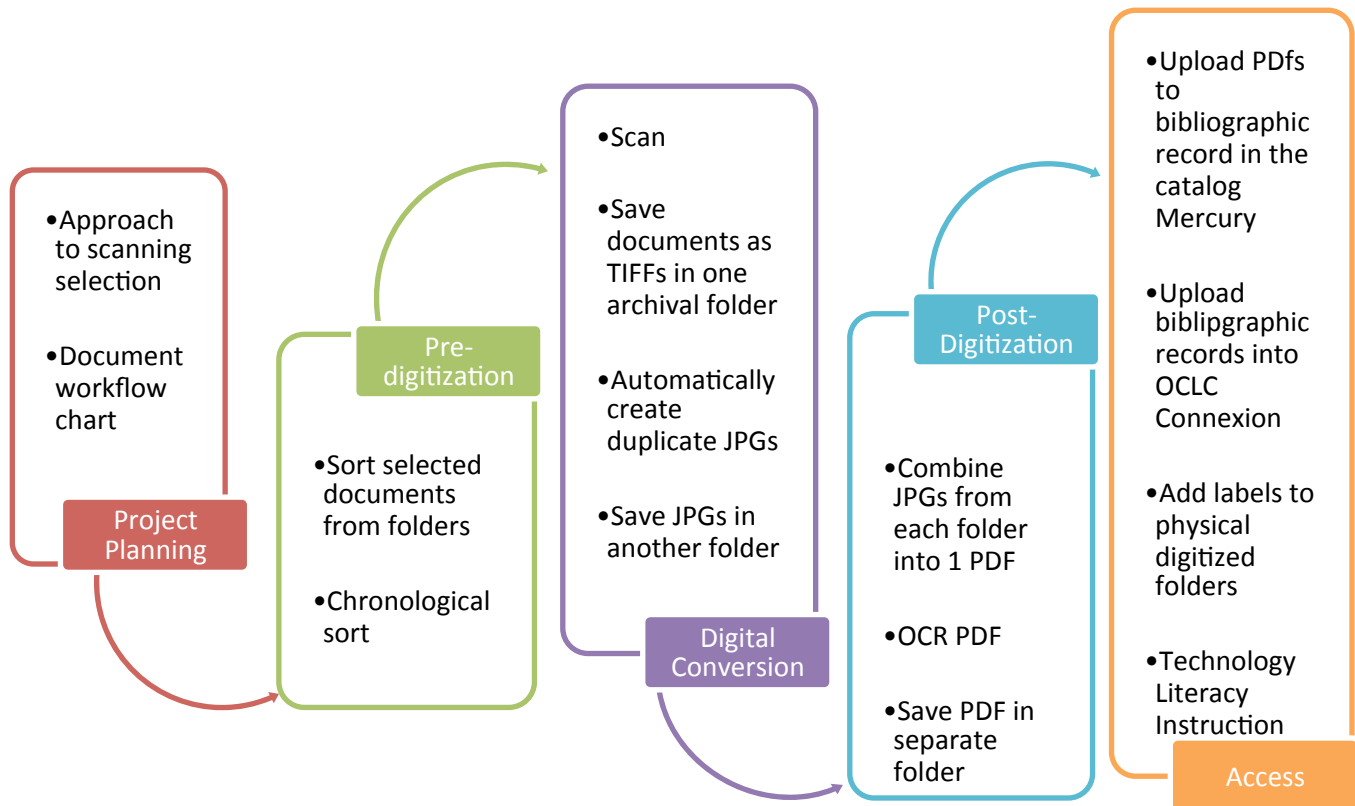
The workflow document will communicate the current status of the digitization of folders as well as the details about the stages a folder took to reach any point in the process, providing a detailed breakdown of its history. This is particularly important when tracking any issues with individual files or issues across the project. A template should be used to track the stages in Microsoft Excel (**See example Table 1 p. 24**).

The following header fields are recommended:

Folder name	Note the name of the folder.
File name	Record the assigned file name. ( <b>See File Naming, p.17</b> )
Date scanned	Track the date the object was scanned.
Type	In this field, only indicate if it is oversized.
Scanned by	Record the name of who scanned the item.
Scanner model	Identify which scanner was used.
Scanning method	Distinguishes between whether scanned or shot with a camera.
Tiff size range (MB)	Track the TIFF file size range of the folder objects.
JPG size range (MB)	Track the JPG file size range of the folder objects.
PDF size (MB)	Note the final combined PDF size.
Bib record #	Number of the bibliographic record the item belongs to. This can be found in the 001 field of the MARC record in the Mercury catalog.
Notes	Add any notes or issues pertaining to the item or project.

The workflow of the project should include 5 phases.

**Figure 1. 5 Phase Project Management Workflow**



## Digitization General Principles

The following are general principles to be followed through the digitization process:

- Document the digitization workflow.
- Achieve the best quality scans within the projected use and size.
- **[Approach 1]**
  - Scan full images without compression as TIFFs Preservation Masters (PM) for preservation.
  - All PMs will be stored on a server separate from the JPGs Surrogates.
  - JPGs will be created from TIFFs.
  - Combined PDF/As will be created from JPGs.
- **[Approach 2]**
  - Direct scans will be created and saved as PDF/As.

## Digitization Workflow

Below is a recommended workflow to accomplish the scanning of physical materials proposed of the vertical files for both Approach 1 and 2.

Figure 2. Recommended Scanning Process for Approach 1

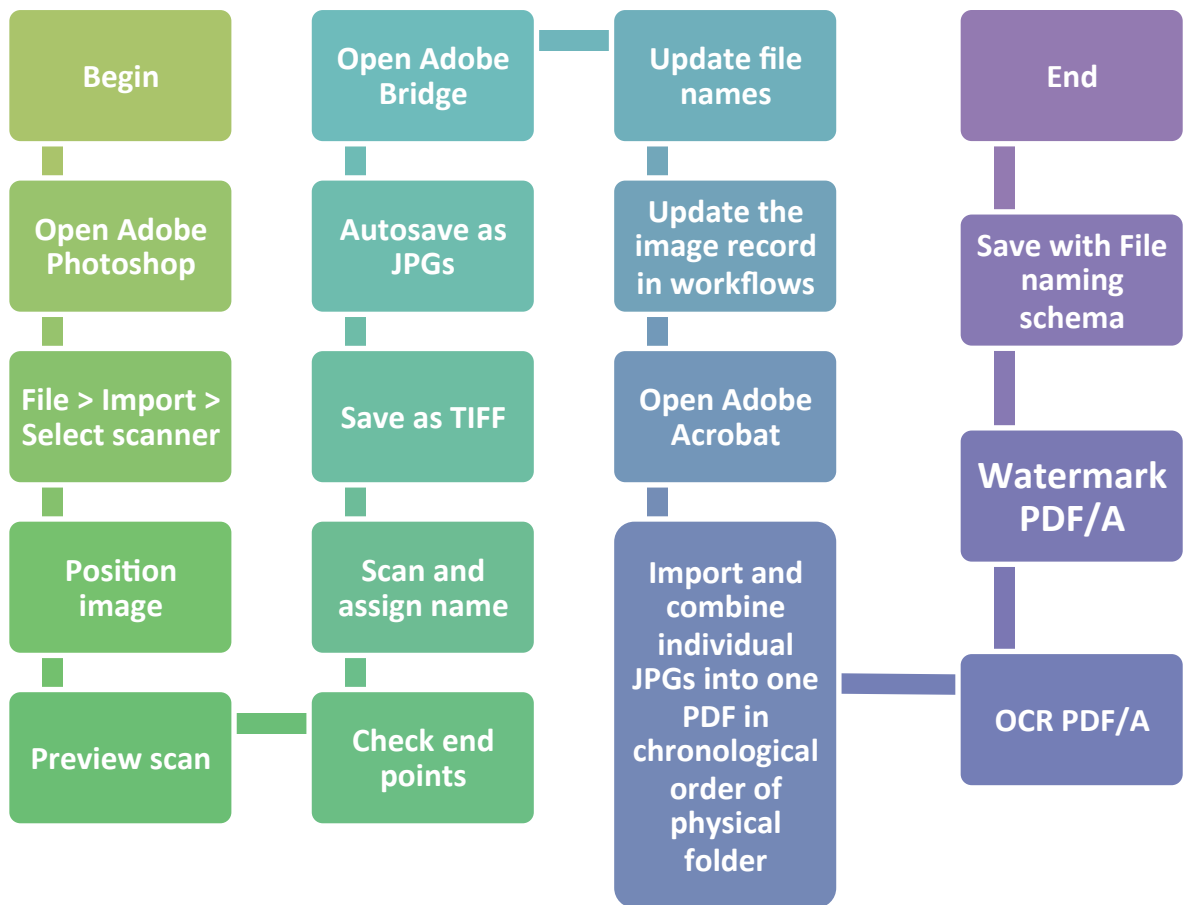
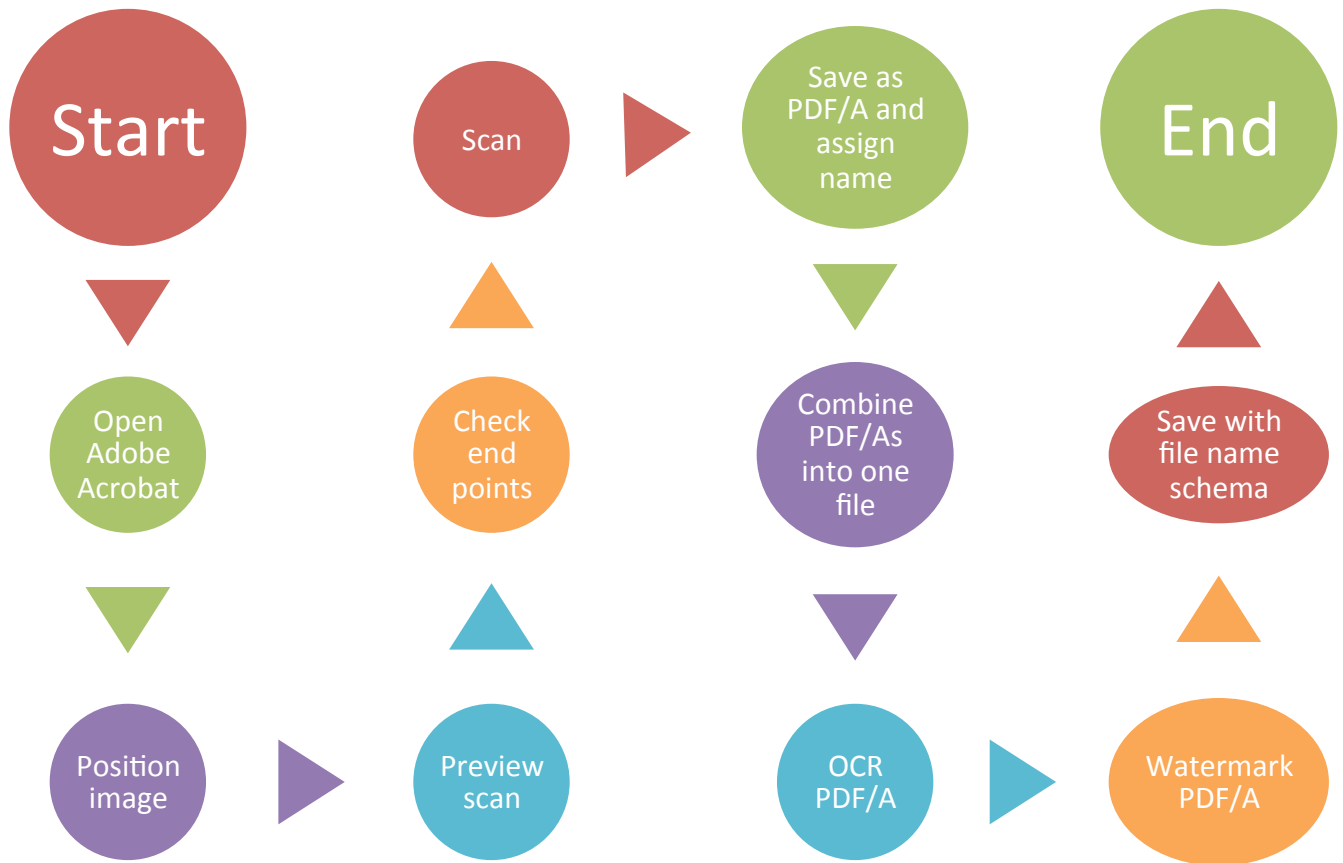


Figure 3. Recommended Scanning Process for Approach 2



# General Digitization Requirements

## Approach 1

- Each original object scanned will have an uncompressed preservation master (PM) and compressed JPG.

## Both Approaches

- Thumbnail images will be generated automatically by DigiTool. if this DAM is still being use.
- In the instance there is no Digital Asset Management System (DAMS), consult with the Information Technology department about strategies for restricted access.

## Digital Asset Management System (DAM)

There is currently no DAM in place in NGA or its library; however this has been discussed and plans are in place to acquire one. Once this is established, it is recommended to find, manage, control, and distribute digital assets and use it to as a central repository for files. Depending on the DAM, access levels can be controlled by file or batch upload in the catalog.

## File Formats and Bit Depth

### Approach 1

- PMs in color are uncompressed TIFF format with 30 bits per pixel and scanned as RGB.
- PMs in grayscale are uncompressed TIFF format with 8 bits per pixel in grayscale and scanned as grayscale.
- Files in color are compressed JPGs.
- Files in grayscale are compressed JPGs.
- Combined PDF/As can include both compressed grayscale and color converted JPGs.

## Resolution and Sizing

- All materials must be scanned at 100%.

### Approach 1

- TIFFs in color should be scanned as RGB at 600dpi.
- TIFFs in grayscale should be scanned at 300dpi.

## Color Management

For monitor calibration see the instruction in the [2009 Digitisation Workflow and Guidelines by Ahmed Abu-Zayed](#). For color correction for the scanner, see the scanner manual or consult your tech support.

## Technical Requirements

Scanning with the first scanner is recommended, however, scanning with the second is sufficient.

1. Konica Minolta PS 7000 Flatbed scanner
2. imageRUNNER ADVANCE C5045

## Software Needed

- Access to the Catalog (Voyager)
- Adobe Acrobat Pro
- Adobe Bridge
- Adobe Photoshop CS3
- Microsoft Excel
- OCLC Connexion

## Calculating Server Space

Server(s) will be a necessity to store the files. Preferably, store the PMs on an offsite server and the JPGs and PDFs on an onsite or server. In order to store 12,000 folders of scanned material, a low average estimation would be 14,000 GB of space is recommended.

You can calculate the space needed by adding the file sizes of TIFFS, JPGs and PDF/As together. Remember every 1000MBs = 1 GB.

Example: TIFFs + JPGS + PDF/As = server space needed

## Metadata

NGA is aware image metadata collection is a best practice of digitization projects, but has decided to not collect it. We have made this strategic decision because this is meant for an internal system and is not projected to be shared beyond the institution walls.

## File Naming

Each digital object must have a unique digital object identifier (DOI). The VFs folder reference will become the DOI. This will include the last name, first name, years and the number of the object scanned in sequence.

Examples:

Last name\_first name

Lewitt\_sol\_1

Or

Last name\_first name\_YearYear `

Lewitt\_sol\_19501959\_1

If a year or a range of years are included on the physical vertical file, they should be included in the DOI.

### *Approach 1: TIFFs and JPGs*

Each page of the objects scanned will have individual PMs. In order to identify them, the DOI will be modified to reflect this at the end in sequential order of being scanned.

Examples:

Lewitt\_sol\_19501959\_1

Lewitt\_sol\_19501959\_2

Lewitt\_sol\_19501959\_3

Lewitt\_sol\_19501959\_4

Lewitt\_sol\_19501959\_5

Lewitt\_sol\_19501959\_6

Lewitt\_sol\_19501959\_7

Lewitt\_sol\_19501959\_8

JPGs will reflect the same DOI as the TIFFs they are created from.

## Multiple Pages in a Document

If there is more than one page in an object being scanned such as several pages in a pamphlet or the back of an announcement, add an underscore after the years to indicate the number scanned in the sequence and after that the number of the scan in the single object. For example the front of a press release would be 1 and the back would be 2 and any scan after would follow.

Last name\_first name\_YearYear\_Object number\_ number of scan in particular object

Last name\_first name\_YearYear\_1

Lewitt\_sol\_19501959\_1



Last name_first name_YearYear_1_2	Lewitt_sol_19501959_1_2
Last name_first name_YearYear_1_3	Lewitt_sol_19501959_1_3
Last name_first name_YearYear_1_4	Lewitt_sol_19501959_1_4
Last name_first name_YearYear_2	Lewitt_sol_19501959_2
Last name_first name_YearYear_2_2	Lewitt_sol_19501959_2_2
Last name_first name_YearYear_2_3	Lewitt_sol_19501959_2_3
Last name_first name_YearYear_2_4	Lewitt_sol_19501959_2_4

## PDF/As

When the JPGs of a folder are combined and converted to a single PDF/A or individual PDF/As are combined, the file naming changes slightly. If there is only one folder, it will be the same as a single TIFF structure, which includes last name and first name as well as dates if applicable.

Examples:

Lewitt\_sol\_19501959

## Future Changes

If a physical folder is changed such as more materials are added or folders are split, the digital files should be updated accordingly. For example, if Sol LeWitt 1950-1959 physical folder grows and needs to be split up and renamed Sol LeWitt 1950-1954 and the other Sol LeWitt 1955-1959, the TIFF, JPG file names must be updated to reflect the change in folder location. The PDF/As must also be reorganized accordingly and the file names updated.

PDF/A file name change example if file split in two:

Lewitt_sol_19501959	>>>	Lewitt_sol_19501954
		Lewitt_sol_19541959

Alter TIFF and JPG files if they change location in physical files.

Example if change folders:

Lewitt_sol_19501959_1	>>>	Lewitt_sol_19501954_1
Lewitt_sol_19501959_2	>>>	Lewitt_sol_19501954_2
Lewitt_sol_19501959_3	>>>	Lewitt_sol_19551959_3

## Creating Preservation Masters TIFF

In Approach 1, each item in a folder will have a PM in uncompressed TIFF format and a JPG will be created in the batch process.

Color RGB is required for colorful or handwritten materials and grayscale should be used for items that are in black and white, grayscale or have tonal variations. There is to be no sharpening, stitching or color enhancements to TIFF files. Originals shall be scanned as they are.

## Cataloging Addition

Add the 583 – Action Note field to the file’s MARC record. Since there is no MARC field that describes attached PDF/A, add the information here.

583 The collection was scanned in the order in which it was filed in the original folder. The entire collection was scanned in [INSERT YEAR].

If funding is acquired for the project use the example below:

583 The collection was scanned in the order in which it was filed in the original folder. The entire collection was scanned in [INSERT YEAR] with funding provided by [INSERT SOURCE OF FUNDING].

# Image Manipulation Guidelines

Minimal image manipulation is recommended.

## Watermarks

Add the NGA symbol to PDF/As only after the process of combination has been completed. The watermark should be faint enough that the researcher or curator can read the text or see the images in the document, but can also see the watermark. This is intended to prevent publishing of the materials from the PDF/As.

## Orientation

- All images should be scanned and saved in its original orientation whether landscape or portrait.
- In regards to documents with multiple pages bound, they should be scanned as a layout in one saved file.

## Cropping

- In no event should the actual document be cropped. If it is too large for the scanner, than a photograph is to be taken of it.
- Whenever possible, a border zone is recommended of one centimeter or less of the background of the scanned document. While this may not always be possible with the scanning equipment and size of the document, one should attempt to capture the margin on all four sides of the document. This assures to researchers the whole document is presented before them.

## Blank Pages

- In the event there is a blank page in a spread of a bound document on one or the other side of text or an image, scan both pages.
- If there is a blank page with no text on one or the other side or it is a single blank page, do not scan it.
- The decision to not scan the single blank pages can be noted in the bibliographic record.

## Copyright

A strategic copyright policy ensures that the library remains true to its mission and provides dynamic access to digitized collections, while operating within the law. A disclaimer is recommended to pop up before the researcher or curator can download or view the PDF/A folder stipulating the use for research only and all materials are not for publication as they are under copyright. This can be worked on with someone from IT. Ideally, a window may pop up before the user can download or view the materials explaining copyright use and getting a digitally signed agreement from the user or it can be mentioned in the bibliographic record.

## Access

Providing new ways of discovery to the collection is important. Adding links to PDF/As to the bibliographic records will enhance access to the VF collection. Uploading the VF bibliographic records into OCLC Connexion is an additional step that makes digitized content available in WorldCat. This will provide another point of discovery and access to the collection on an international level not previously implemented before.

## Technology Literacy

Researchers and curator's levels of familiarity with technology may vary. An information literacy class teaches users how to find, download, and view the PDF/As in a browser. Also, since the PDF/As are OCRed, instructors should teach curators and researchers how to search downloaded content or browse PDF/As as if they were browsing the physical folder item by item in chronological order.

## References

- (2010). Artist Files Revealed. ARLIS/NA. Retrieved from <http://www.artistfilesrevealed.com/tiki/tiki-index.php>
- (2010). Creating a Digital Smithsonian: Digitization Strategic Plan. *Smithsonian Institution*. Retrieved from [https://www.si.edu/content/pdf/about/2010\\_SI\\_Digitization\\_Plan.pdf](https://www.si.edu/content/pdf/about/2010_SI_Digitization_Plan.pdf)
- (2013). Wikipedia:GLAM/Israel Museum, Jerusalem/Israeli Artist Files. Wikipedia. Retrieved from [http://en.wikipedia.org/wiki/Wikipedia:GLAM/Israel\\_Museum,\\_Jerusalem/Israeli\\_Artist\\_Files](http://en.wikipedia.org/wiki/Wikipedia:GLAM/Israel_Museum,_Jerusalem/Israeli_Artist_Files)
- Archives of American Art (2013). Cataloging Guidelines for Collections. Fully Processed and/or Fully Digitized Collections. *Archives of American Art*. Retrieved from <http://www.aaa.si.edu/collections/documentation/cataloging>
- Archives of American Art (2013). Technical Documentation Appendix B: Examples of Handling and Special Problems for Scanning Technicians. *Archives of American Art*. Retrieved from [http://www.aaa.si.edu/files/forms/ForScanningTechnicians\\_ArchivesofAmericanArt.pdf](http://www.aaa.si.edu/files/forms/ForScanningTechnicians_ArchivesofAmericanArt.pdf)
- Abu-Zayed, Ahmed (2009). Digitisation Workflow and Guidelines. *University of Exeter*. Retrieved from <http://projects.exeter.ac.uk/charter/documents/DigitisationWorkflowGuidev5.pdf>
- Ashenfelder, Mike (2013). You Say You Want a Resolution: How Much DPI/PPI is Too Much? *Library of Congress*. Retrieved from [http://blogs.loc.gov/digitalpreservation/2013/07/you-say-you-want-a-resolution-how-many-dippi-is-too-much/?goback=.gde\\_52874\\_member\\_259685208](http://blogs.loc.gov/digitalpreservation/2013/07/you-say-you-want-a-resolution-how-many-dippi-is-too-much/?goback=.gde_52874_member_259685208)
- Erway, Rick (2011). Rapid Capture: Faster Throughput in Digitization of Special Collections. *OCLC Research*. Retrieved from <http://www.oclc.org/content/dam/research/publications/library/2011/2011-04.pdf?urlm=162955>

Federal Agencies Digitization Guidelines Initiative-Still Image Working Group (2009). Digitization Activities Project Planning and Management Outline. *Federal Agencies Digitization Guidelines Initiative*. Retrieved from <http://www.digitizationguidelines.gov/guidelines/DigActivities-FADGI-v1-20091104.pdf>

National Gallery of Art. (2010). IT 2020:IT Strategic Plan 2010-2020. *National Gallery of Art*. Retrieved from [http://inside.nga.gov/tdp/pdf/2011\\_it\\_strategic\\_plan\\_09\\_18\\_09.pdf](http://inside.nga.gov/tdp/pdf/2011_it_strategic_plan_09_18_09.pdf)

Solomon R. Guggenheim Museum. Library Department (2013). Solomon R. Guggenheim Museum Archives. Retrieved from <http://www.guggenheim.org/new-york/collections/library-and-archives/archive-collections/A0008/#>