

Finding Aids: PDF/A Standards for Digital Reproductions

Introduction

PDF/A-1b files can be used to deliver web-accessible reproductions of transcripts, translations, indexes, or other scanned collection content. When making PDF/A-1b files, use Adobe Pro to add OCR, descriptive metadata and copyright information, and to edit bookmarks.

Paper objects may be digitized and assembled as PDF files using a camera, a scanner, or the printer/photocopier, and by further processing the files in Adobe Pro. Word-processed files, power-point presentations, or spreadsheets can be saved as PDF files in Microsoft Office and further processed in Adobe Pro.

General Principles

Four general principles govern the creation of PDF/A files:

1. OCR

Full-text, keyword searchability adds value. Because optical character recognition (OCR) enables full-text searches as well as voice recognition readers for the visually impaired, OCR can be added to most files containing typewritten text. Files that are primarily hand-written should not be subjected to OCR.

Add OCR to enable text searching of files that contain images of typewritten text only. Files where the text is obscured by annotation, is faint, illegible, or is primarily hand-written should be considered on a case-by-case basis. OCR does not need to be added to MSOffice files (or other spreadsheets or word-processed documents) that are saved as PDF files.

2. Document Properties (Metadata)

Added metadata should identify the original characteristics of a reproduction, provide a citation to its parent collection and repository, and provide generic information indicating that the material may be protected by copyright.

Document properties that describe the creator, title, and source of an archival component are always to be added to the Extensible Metadata Platform (XMP) header of any PDF/A file. XMP is a labeling technology that allows us to embed data about a file into the file itself. Data that we typically record includes descriptive metadata naming the creator and title of the original collection component, the title of the collection and Minnesota Historical Society as the holding repository. As well as providing descriptive information about the file, these properties are essential to web indexing performed by search engine crawlers.

3. File Naming and File Structure

File names and file folders for digital reproductions should follow consistent practice. To determine how to structure folders and name files, consider whether or not master files are stored on the digital preservation server. When master files are preserved, folders and filenames for access versions should be retained to mimic those stored on the preservation server. For information regarding file naming and file structure, see the EAD Tagging Guidelines.

4. Bookmarks

Bookmarks that delineate meaningful parts within a PDF file are like a table of contents that provide useful navigation to specific content within the file.

Processors should pay attention to bookmarks and make decisions about whether to retain, eliminate or edit bookmarks and how they should be named.

Retain bookmarks that were made by combining multiple files into one PDF file when the bookmarks parallel the names of master files stored on the preservation server. These bookmarks identify the name of each master file contained within the PDF copy.

Retain bookmarks that provide meaningful section breaks and navigational links such as the titles of letters within a correspondence file, the titles of photographs within an album, or the titles of chapters and appendices within a report. These kinds of bookmarks can help people more quickly browse the contents of a file.

Bookmarks that do not provide meaningful breaks or navigational links should be deleted or revised in a manner that imparts meaning. These types of bookmarks may be no more than a series of camera image or preliminary accession numbers assigned to JPG or TIF files as items were scanned.

Metadata Standards

For each PDF/A file, follow the metadata standards detailed below when creating or preparing a file for web access.

Document Properties

Add XMP metadata to the Document Properties for each PDF/A file that describe the content of the file. Record title and author data that describes the original component that has been digitized.

Title

Mandatory

Enter the title of the item, folder, or component that has been scanned. This title should describe the component in the same manner it is titled in the corresponding inventory. When needed to distinguish one folder from another, folder numbers can be added. Include the date of the component.

Examples:

Jimmy Carter to Walter F. Mondale, February 16, 1978
Foreign Policy Breakfasts, 1977-1978, folder 2
VP Role and functions, 1976-1977

Author

Mandatory

Enter the specific creator of the content that has been digitized. If the content represents a single item, give the name of the creator of the item. If the content represents an entire folder, give the name of the folder's creator. Use the LCNAF or MnPALS authorized form of a name, except record personal names in direct order: first, middle (initial), last name. Include birth and death dates as recorded in authorized forms, except enclose the birth and death dates in parenthesis.

This field should also be used to indicate additional creators (such as a translator or secondary author). Use a semi-colon to distinguish one creator from another. A relater term may be included when necessary to clarify the responsibility a named author had in producing the work.

Examples:

United States. President (1977-1981 : Carter)
Walter F. Mondale (1928-)
Charles John LaVine, Translator
Great Northern Railway Company (U.S.)
Minnesota. Department of Education
Minnesota. Governor (1999-2003 : Ventura)

Subject (Description)

Optional

Enter a very brief description of the item or folder. This can be used to highlight relevant dates, and/or any remarkable content.

Examples:

Agendas, briefing materials, and handwritten notes regarding foreign policy.
Includes draft of United Nations Charter dated October 24, 1985.

Keywords

Mandatory

Enter the series and collection title from which the content was scanned and add "Minnesota Historical Society." This information is provided in the Overview section of an EAD inventory.

Example:

Vice Presidential Papers. Walter F. Mondale Papers. Minnesota Historical Society.

Correspondence of Walter Mondale's Office. Attorney General. State Archives. Minnesota Historical Society.

Steam Locomotive Photographs. Mechanical Department. Great Northern Railway records. Minnesota Historical Society.

Copyright

Copyright

Required

The pull down bar gives you three options, "Unknown," "Copyrighted," and "Public Domain." Select and enter "Unknown."

Copyright Notice

Required

Enter "This material may be protected by copyright law (U.S. Code, Title 17)."

Copyright URL

Required

Enter "<http://www.mnhs.org/copyright>" in the Copyright Info URL to provide an external link to additional copyright information.