Finding Aids: PDF Standards for Legacy Inventories

Introduction

PDF versions of legacy inventories can be a highly economical alternative for producing web accessible versions of finding aids in cases where an old inventory cannot easily be migrated into EAD. In addition to being a low cost method, the PDF file format is highly versatile and can be used for typescript inventories that exist only on paper as well as for inventories in word processed or spreadsheet files.

Typescripts can be scanned to either PDF or Word file formats using the suite's copier/printer/scanner. Word processed and spreadsheet files can be saved as PDF files either directly in MS Office applications or by importing the files into Adobe Acrobat. If a revision to a typescript is needed, a PDF file can be scanned, exported to Word using Adobe Acrobat, edited in Word, saved in a Word format, and then saved again as a PDF file.

Due to a scanner's ability to correctly interpret typescript characters and the variety of legacy inventories you may encounter, you may wish to experiment with any single document to determine whether a PDF scan, a Word scan, or a PDF to Word export is the most advantageous method.

When revisiting a typescript inventory we generally want to create a PDF version for public access if the typescript version is neat and legible, with no handwritten changes, additions, or annotations.

Please note that PDF inventories are not indexed by solr in Collection Finding Aids, but they are indexed on the MNHS website and by general web search engines. Because XML data will serve us longer than Word or PDF files, EAD inventories are preferred, but PDF versions can be an acceptable alternative.

Follow the general principle and standards below when producing PDF versions of legacy inventories.

General Principles

Informational Content

Make a PDF only of inventories with substantive information, folder titles, or file lists. When an inventory is not sufficiently detailed, do not create a PDF version, but consider eliminating the inventory by adding box contents or other notes to the MARC record.

Legibility

When revisiting a typescript inventory, we generally want to create a PDF version for public access only if the typescript is neat and legible with no handwritten changes, additions, or annotations.

OCR

Because optical character recognition (OCR) enables text searching as well as audible readers for the visually impaired, OCR is to be added to typescript inventories.

Document Properties

Document properties that describe the title, author, and subject of a file are always to be added to any PDF file. These properties are essential to web indexing performed by search engine crawlers (Google, etc.).

Copyright

Because PDF versions of MNHS finding aids are published on the MNHS website, they are understood to be copyrighted by the Minnesota Historical Society, so we will not add copyright properties nor attach a copyright page to the PDF.

Standards

OCR

Add OCR to enable full-text searching of all inventories that are scanned from paper typescripts.   OCR does not need to be added to word processed files or spreadsheets that are saved as PDF files.

Document Properties

Document properties are used by a database application that builds the index pages to our finding aids as well as by search engine crawlers.

Add document properties that describe the inventory by identifying the title of the inventory, the author of the inventory, and a brief abstract that describes the collection.

Title

The title property identifies the title of the inventory and is used by our finding aids A-Z index to list the title of each inventory.

Transcribe the title of the inventory or construct a title that describes the finding aid from the associated MARC record. The title given in the PDF file should follow the standard for EAD finding aids and should include the name of the creator followed by the title of the inventory.

Examples:

Secretary of State: An Inventory of Its Land Survey Plats at the Minnesota Historical Society

David Durenberger: An Inventory of His Papers at the Minnesota Historical Society

Author

The author property should reflect the author of the inventory. Because our collection inventories represent a collaboration of staff efforts, enter "Minnesota Historical Society" for inventories created by MNHS staff rather than enter individual staff names.

For scans of inventories that may have been created by another agency, such as microfilm pamphlets that accompany NARA microfilm publications, enter the name of the agency or author that appears on the inventory.

Examples:
Minnesota Historical Society
National Archives and Records Administration (U.S.)

Subject

Copy and paste the abstract from the inventory or MARC summary note. For MARC records that have extremely long summary notes, edit the summary to enter a subject property of 1-2 sentences in length.

Example:
The record set of state agency administrative rules and regulations, as issued by the reviser of statutes and filed with the Secretary of State.

File Formats

Legacy inventories that are to be published on the finding aids website and linked from catalog records should be saved in PDF format. The PDF/A file format is not needed for legacy inventories. In instances where there is already an electronic Word inventory we typically will create a PDF, and maintain both the Word and PDF versions.

File Naming

PDF files made from word processed files are assigned filenames that correspond to the filename of the existing Word inventory (e.g., 00259.docx; 00259.pdf). When a PDF version is made from a typescript inventory (where there is no existing inventory in electronic form) assign a filename in the usual manner and record the name in the Finding Aid File Names list.