

## MOUNT HOLYOKE COLLEGE ARCHIVES AND SPECIAL COLLECTIONS GUIDELINES FOR ELECTRONIC RECORDS FORMATS

This document describes electronic records formats that are supported by LITS. Using supported file formats improves the likelihood that electronic records will remain accessible and usable in the future. This document also describes how to convert a document to the PDF/A format, the recommended file format for archiving text documents. The last part of this document includes recommendations for naming the electronic files.

### FILE FORMATS

Electronic records must be managed to insure authenticity, integrity as well as discoverability over time. Though the proprietary nature of many file types makes it impossible to make guarantees, the following are formats that LITS currently supports:

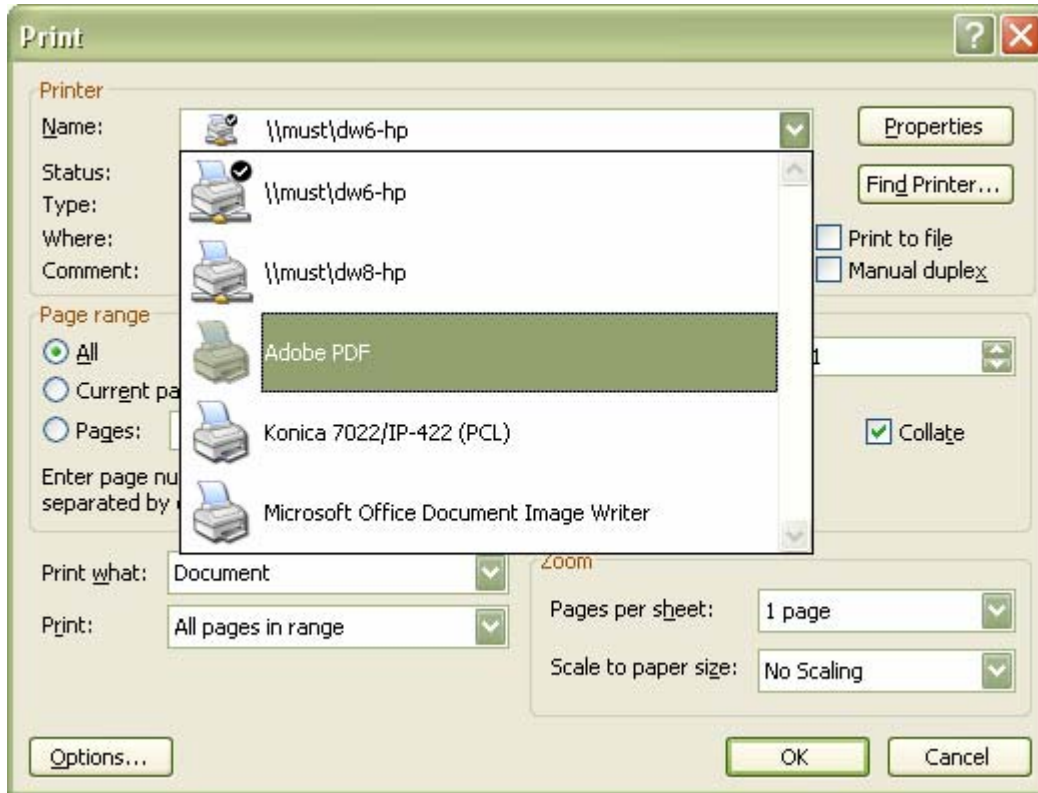
MIME type	Description	File Extension
Text	PDF (Portable Document Format)	.pdf, pdf/a
Audio	AIFF (Audio Interchange File Format)	.aiff, .aif, .aifc
	WAV (Waveform Audio Format)	.wav
Image	JPEG (Joint Photographic Experts Group)	.jpeg, .jpg
	TIFF (Tagged Image File Format)	.tiff, .tif
Video	MPEG (Moving Pictures Expert Group)	.mpeg, .mpg, .mpe

All electronic College records should be stored on the network so that they are backed up and retrievable by officials of the College. Electronic documents can also be printed and stored in any paper-filing system for administrative use and later transferred to Archives and Special Collections, when a record has reached the end of its life cycle.

To further protect electronic documents for long-term preservation, all text files having permanent value should be saved as a pdf/a file. To determine if a record(s) has permanent value, please email [archives@mtholyoke.edu](mailto:archives@mtholyoke.edu) or telephone (413) 538-3079.

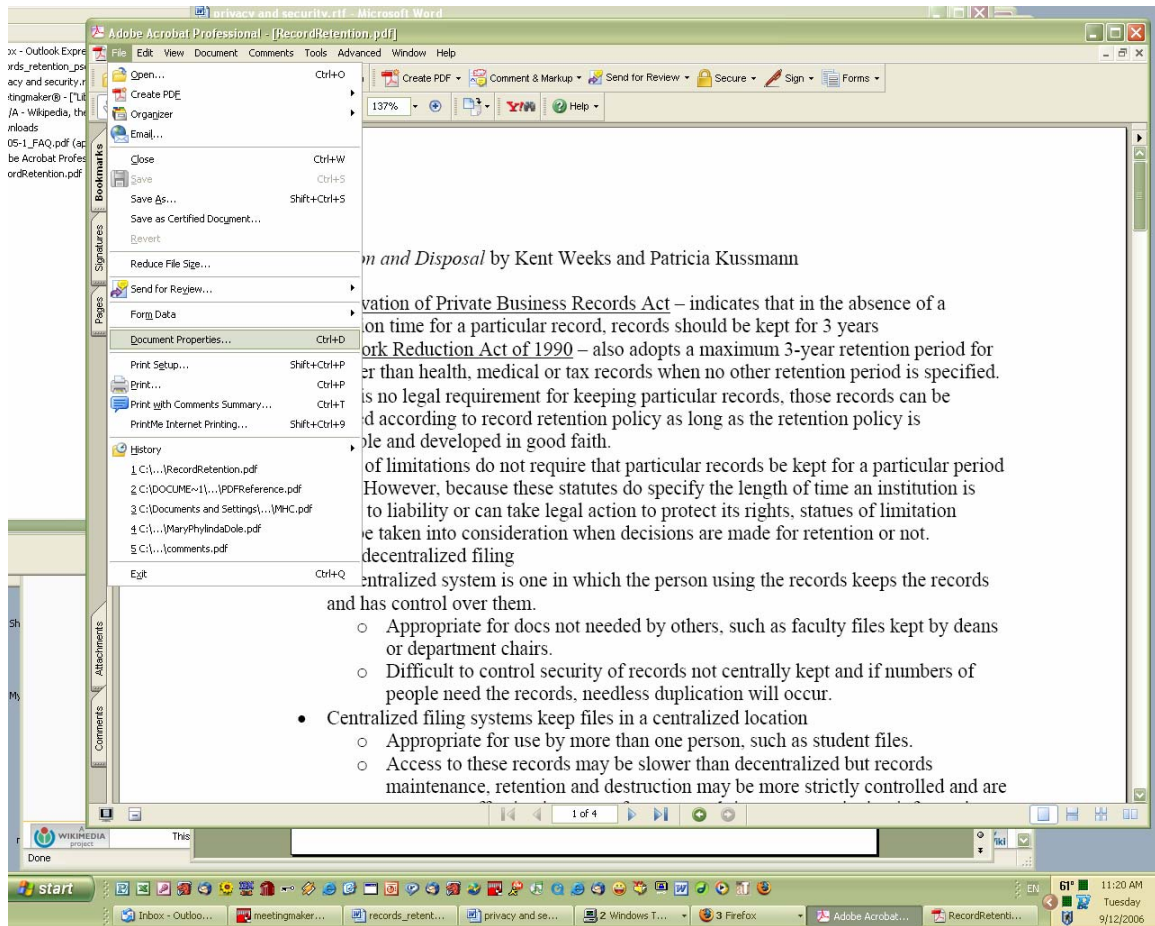
### PDF-A: WHAT IS IT AND HOW DO I MAKE ONE?

Most of us have created or used PDF files in the normal course of our work. Though Adobe is proprietary which means the file can only be read using for-profit software, it is well-documented and information professionals are confident that the files will be accessible for the long-term regardless if Adobe were to, for example, go out of business. For archiving purposes, however, the PDF needs to conform to particular rules to be considered archival. PDF-A (archival) disallows or limits features that could complicate long-term preservation. Please see appendix for more detailed information about the standard.



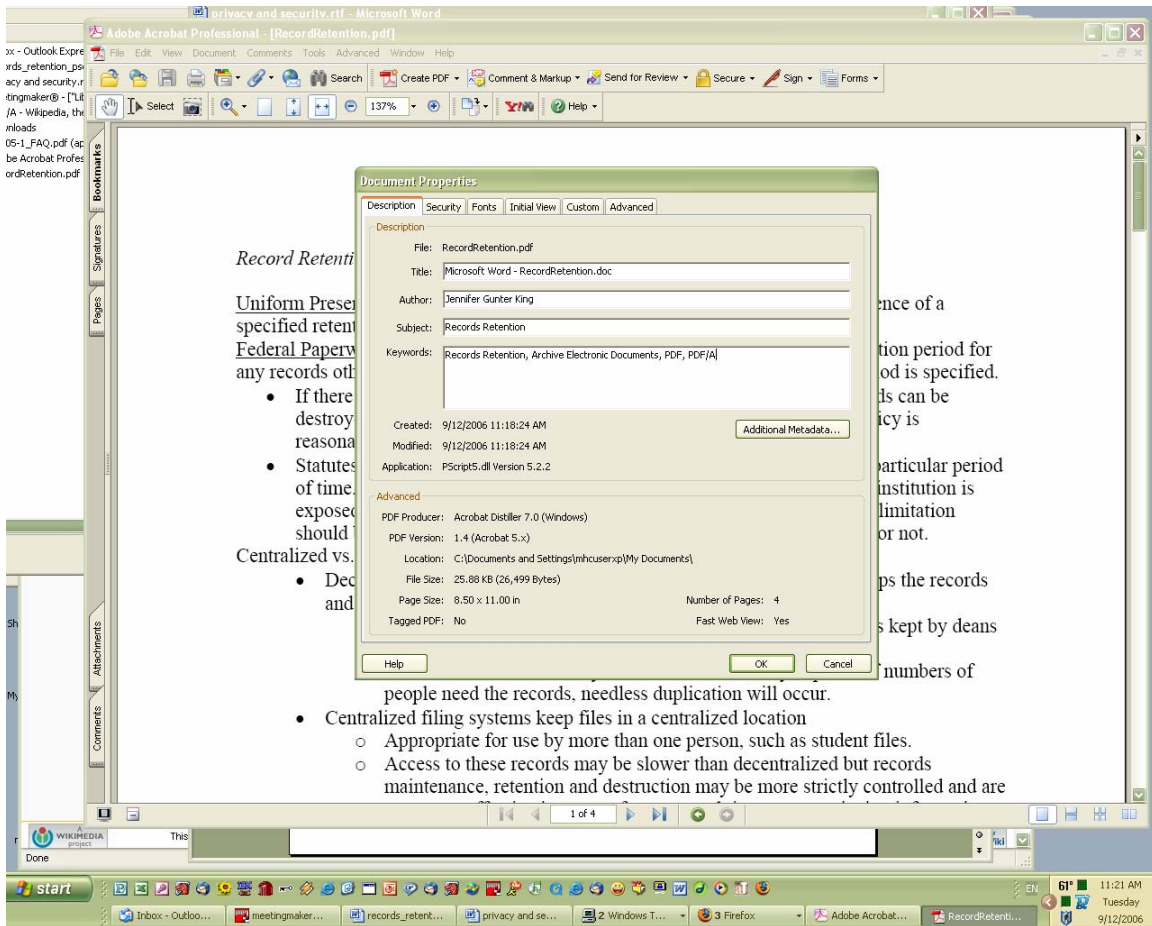
### Creating a PDF-A Document

You must have Adobe Acrobat 7.0 Professional installed on your computer. From Microsoft Word or other office program, select “File” “Print”. When the print screen appears, select “Adobe” as printer and then “ok”. This will convert your document to the .pdf format.



## Editing document metadata

Choose File>Document Properties and in the description tab enter pertinent information. There are also advanced fields. In the same Document Properties click “additional metadata”. You can import your own metadata and share metadata code among documents.



Once you have created your pdf/a document, Archives and Special Collections recommends naming the file according to established standards.

**FILE NAMING PROTOCOLS** (Based on the Document Information Dictionary in the ISO 19005-1:2005 Standard [6.7.3] and standards for date and time formatting)

Document names should include the date as well as a key word (or more) followed by a .format.

The international format defined by ISO (ISO 8601) defines a numerical date system as follows: YYYY-MM-DD where

- YYYY is the year [all the digits, i.e. 2012]
- MM is the month [01 (January) to 12 (December)]
- DD is the day [01 to 31]

For example, "3rd of April 2002", in this international format is written: 2002-04-03.

Example of file naming:

20060602\_seniorstaff.pdf  
20060602\_opc\_minutes.pdf

## SOURCES FOR FURTHER INFORMATION

Document Management—Electronic document file format for long-term preservation  
ISO 19005-1: 2005

NARA Document for PDF Records (both compliant to PDF-A and their own specifications)  
<http://www.archives.gov/records-mgmt/initiatives/pdf-records.html>

AIIM: PDF-A Fact Sheet, Standards  
[http://www.aiim.org/documents/standards/19005-1\\_FAQ.pdf](http://www.aiim.org/documents/standards/19005-1_FAQ.pdf)  
<http://www.aiim.org/standards.asp?ID=25013>

Adobe XMP  
<http://partners.adobe.com/public/developer/xmp/topic.html>

Editing Document Metadata: Adobe Acrobat 7.0 Professional Help File

## APPENDIX: CHARACTERISTICS OF PDF-A

Archival Standard for Documents (ISO 19005-1:2005) maximizes:

- Device independence (consistently rendered independent for platforms)
- Self-Contained (contains all resources necessary for rendering)
- Self-Documenting (contains its own description)

Attributes of a PDF/A Document:

- Two Levels of Conformance:
  - Level A (tagged PDF, UNICODE Mapping)
  - Level B (not tagged)
- Uniform File Format (header, trailer, no encryption)
- Device-independent rendering of graphics
- Embedded fonts, character encoding (see below)
  - What are Embedded Fonts? What are 14 Fonts? Permitted fonts include: Courier (Regular, Bold, Italic, and Bold Italic), Arial MT (Regular, Bold, Oblique, and Bold Oblique), Times New Roman PS MT (Roman, Bold, Italic, and Bold Italic), Symbol, and ZapfDingbats.
- Annotations restricted, content should be displayed by readers
- *External actions restricted*, no dependence on external content
  - Actions are *Launch, Sound, Movie, Reset Form, Import Data, and Java Script*. These items are not allowed inside the PDF. [6.6.1]
  - No hypertext links are allowed unless they are rendered “non-actionable” [6.6.3]
- Readers not required to act on hyperlinks
- XMP metadata “Adobe XML Metadata Framework”  
XMP [Extensible Metadata Platform]  
Adobe PDF documents created in Acrobat 5.0 or later contain document metadata in XML format. XMP provides Adobe applications with a common XML framework that standardizes the creation, processing, and interchange of document metadata across publishing workflows. The Metadata includes information about the document and its contents (author’s name, keywords, copyright information that can be used by search utilities. The document metadata contains, but is not limited to, information that also appears in the Description tab of the Document Properties dialog box.
- No OCR -- The guidelines specified by the ISO 19005-1:2005 do not allow for programs designed to convert documents to PDF/A to be OCR’d (Optical Character Recognition). OCR is a lossy process and original data might be lost.

1/29/2007