

OPTIMIZATION OF PDF FILES TO REDUCE FILE SIZE

PDF Files can store digital text/images as well as scanned images. When a document is scanned and stored as a PDF the file size tends to be rather large since the PDF file must contain the image without knowledge of what the image contains. Storing a scanned document as an image also prohibits the ability to perform text searches within the document.

Many scanning hardware and software support the ability to create PDF files. Some of this hardware and software support the ability to “optimize” a scanned PDF document. Typically the optimization will perform optical character recognition (OCR) as well as some image enhancement and cleanup, based on the program settings. This optimization will typically reduce the file size of the PDF and will also make the PDF text searchable.

PDF scanned files that are optimized by conversion software with Optical Character Recognition (OCR) capability (i.e., Adobe Acrobat software):

- are superior in quality to a scanned PDF.
- have reduced file size as compared to the same document saved as a PDF from a scanner. The software converts a document to the smallest file available within the coding parameters of that software.
- **Hyperlinks within PDFs are not interrupted by “file optimization.”**

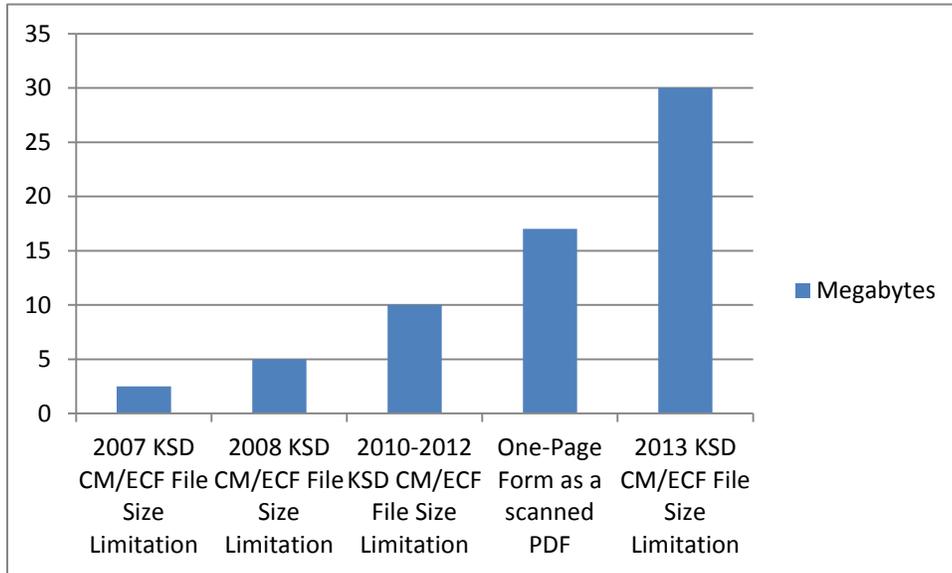
Within Adobe Acrobat, there are two options for “minimizing” space required for PDFs:

- 1) Click “Save As” an “Optimized PDF”
- 2) Under Tools and Document Processing select “Optimize Scanned PDF”

(**Note:** “Compressed files” technically refer to Zip file formats, which are not compatible with CM/ECF.)

PDF files created with scanning equipment and not optimized require the greatest file size.

- KSD CM/ECF has encountered electronically filed documents representing a scanned one-page form totaling **17 MB** of server space. **By comparison, a 50-page text document converted to a PDF in Word resulted in 199 kB or approximately 0.2 MB.** (Currently, the CM/ECF system can accept a document up to 30 MB. There is no stated limitation to the number of documents that can be electronically filed within any given case.)



- The resultant scanned document quality is directly related to the scanning equipment available for use and the equipment’s “dots per inch” (dpi) settings at the time of the scan.
- All pro se filings must be scanned to be placed within the CM/ECF system.
- “File Optimization” of scanned documents to reduce file size may “mis-interpret characters” and produce altered content.



For example, a 6-page document with no graphics was used to determine the range of storage required when saved in various formats.

Format Options	Storage Space Required
Original Word format file (6-pages)	30.2 kB
PDF created with Word	183 kB
PDF created with Scanner	234 kB
“Reduced size” PDF created with Scanner	75.3 kB
“Optimized” PDF created with Scanner	75.3 kB
(All of the PDFs were configured as “searchable”.)	

The file size of any given document scanned to a PDF will be directly impacted by the type and quality of the scanner hardware used and the scanning software settings at the time the scan is accomplished. Greater file sizes of scanned PDFs will be particularly prominent with the reproduction of forms, photos and/or graphics.