

# METADATA

*The District of Maine has compiled the below questions and answer from a Memorandum by Michel Ishakian from the Administrative Office of the Courts dated September 7, 2012 regarding metadata in court documents.*

## FREQUENTLY ASKED QUESTIONS

**Q.** What is Metadata?

**A.** Strictly defined, metadata is data about data. There are two broad views of metadata and both share the common characteristic that it is information associated with a file that does not get printed or displayed. Metadata becomes a problem when it is accidentally released. There are two types of metadata, *revision metadata* and *file description metadata*.

*Revision metadata* is a transitory form of metadata that helps in the editing of a document. The user sees this data when reversing an edit or other changes by using the **undo** command. This command is activated by clicking on the counterclockwise arrow or by typing **Control + Z**. *Revision metadata* remains available until the document is saved using the **save as** command, but is not necessarily removed when using the **save** command.

*File description metadata* helps to manage the file and includes but is not limited to document summary, headers, footers, hyperlinks, Object Linking and Embedding (OLE) information, and routing slip information.

It is important to note that both types of metadata vary among software applications. Problems with metadata being inadvertently revealed are usually associated with documents still being edited. Many applications have a feature to retrieve deleted data (text or numbers) such as the **undo** command or undo/redo history feature. Deleted data is available for use as metadata until the user performs a **save as**, **print**, or **convert** document to a PDF.

**Q.** Between document generating software applications like WordPerfect and Adobe Acrobat, which has the biggest problems with metadata being disclosed accidentally?

**A.** This problem is predominantly associated with documents in an editable format like WordPerfect or Word that have simply been **saved** without taking any additional precautions such as removing hidden data or disabling "Track Changes" in Word and "Save Undo/Redo Items" in WordPerfect.

**Q.** Are files in CM/ECF at some risk of containing hidden metadata?

**A.** Yes, it is possible that some files in CM/ECF may contain hidden metadata. Files in CM/ECF are in PDF format. These final form documents are not where metadata or versioning data has been a problem. Printing files to PDF is one of the ways to remove *revision metadata*. PDF files will still retain some basic *file description metadata*, such as file name and date.

**Q.** What metadata is associated with a WordPerfect document?

**A.** WordPerfect documents include: comments, hidden text, annotations, undo/redo history, document summary data, headers, footers, hyperlinks, OLE object information, and routing slip information.

**Q.** What metadata is associated with a Microsoft Word document?

**A.** Microsoft Word documents include: comments, revision marks from tracked changes, document version information, ink annotations, document properties, (including information from the Summary, Statistics, and custom tabs of the document Properties dialog box), E-mail headers, routing slips, send-for-review information, document server properties, document management policy information, databinding link information for databound fields (last value will be converted to text), user name, template name, and text that is formatted as hidden (a font effect that is available in the font dialog box).

**Q.** What metadata is associated with Adobe Acrobat (PDF) documents?

**A.** PDF documents include: embedded content; attached files; scripts; hidden layers; embedded search indexes; stored form data; review and comment data; hidden data from previous document saves; obscured text with images; comments hidden from the body of the PDF; unreferenced data, links, actions and JavaScript; and overlapping objects.

**Q.** Once I convert a file to PDF, what kind of metadata may still be found and why is it there?

**A.** Creating a PDF using Adobe Acrobat will scrub the document's revision metadata created in the authoring word processor. However, it will retain the file description metadata unless it is properly scrubbed prior to converting to a PDF. This can be done using tools provided within the authoring word processor or other tools available. Some of the file description metadata inherited includes document author, document title, document keyword, etc.

**Q.** How do I remove metadata?

**A.** The process of removing metadata varies depending on the software application being used. For the most current information regarding the removal of metadata from the version of software you are using, you should review the Help section of the software being used.