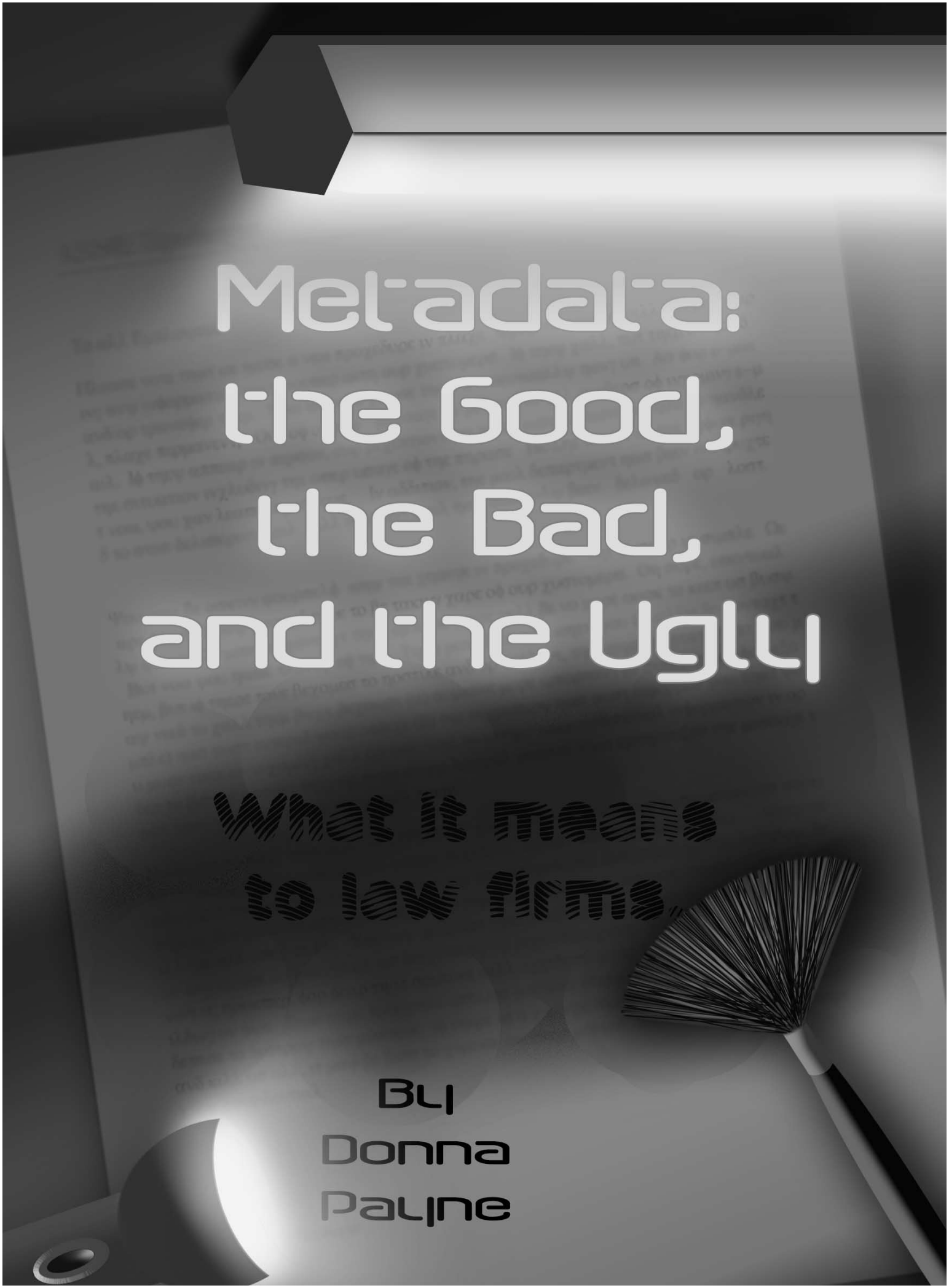

Metadata:
the Good,
the Bad,
and the Ugly
What it means
to law firms
By
Donna
Payne

The other day I was watching a Sunday morning talking heads political show (think forum with four or five political consultants of different persuasions arguing) when a comment made by one of the panelists caught my attention. In discussing a leaked political memo, she queried whether there would be an electronic paper trail to show who had accessed the memo or other information about the journey of the document.

In the legal community, we are all too aware that documents in their electronic format can contain hidden information we call metadata. This information can include when the file was created, the amount of time spent editing, the last 10 authors, and the full filename and path of the computer where these authors made their changes.

Most of the information written about this topic has concerned metadata in Microsoft Word documents. For that reason, this article will tackle not just Microsoft-embedded metadata, but that of Corel WordPerfect and even Adobe Acrobat. Yes, metadata exists in these applications too, as well as other software programs. And in this case, what you don't know can hurt you.

The Good Side of Metadata

The technical definition of metadata is data about data. When viewed, this data provides information about or documentation of information stored within an application or environment. As I type this article, without thinking about it, I am creating some type of metadata. The author metadata is Donna Payne. The publisher metadata is defined as James Publishing: Law Office Computing. Other metadata might include date of creation, date of publication, copyright and more.

If you use a document management system, this should all sound familiar. Document management systems use metadata entered into a profile to assist with searching and retrieving, even purging, certain document types. Profile metadata might include author, client, matter, date created, date modified, e-mail from, e-mail to, keywords and more. Many document management systems have taken this to another level and allow you to set profile information (metadata) to folders. When new documents are created and saved to the folder, the documents automatically take on that folder's metadata.

So metadata isn't a bad thing. It's only bad when it's embedded into a document without your knowledge, and accessible by others who might have malicious intent.

Corel WordPerfect Metadata

Much hoopla has been raised about metadata in Word documents, but WordPerfect falls victim to this exposure as well.

When your document is saved in WordPerfect, the Undo/Redo information might be retained and be accessible to others. This includes all of your document's history regarding text that was cut, copied and even deleted.

Undo/Redo History. If you want to see the last 300 actions of a WordPerfect document, just change the "Undo/Redo History" setting. When you send the document, choose to save this history with the document, and send the file electronically to someone else to edit. When the document is returned, you can track, accept or reject changes made to the file.

This feature is designed to allow you to undo a series of actions, or to undo all edits at once. WordPerfect also keeps a temporary file for each undo level, so if you are worried about discovery, don't forget about these saved temporary files on your computer.

Removing Undo/Redo History. From the "Edit" menu, choose "Undo/Redo History." Figure 1 is an example of the stored Undo/Redo stack in the document.

Figure 1. Undo/Redo History.

1. Click "Options." The "Options" dialog box is displayed. Note the "Save Undo/Redo items with document" option is enabled and the "Number of Undo/Redo items" is set to 300 (see Figure 2).
2. Remove the check mark next to "Save Undo/Redo items with document" and click "OK" then "Close."

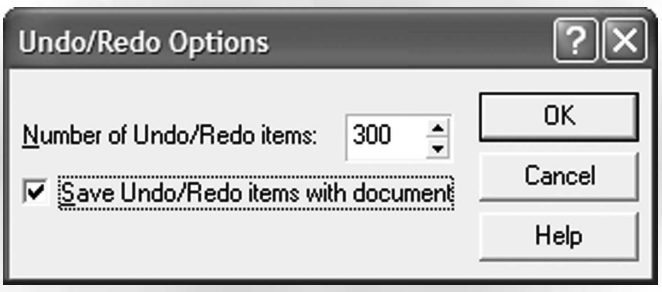


Figure 2. Undo/Redo Options.

3. Now it's important to save the document to remove the saved list of items.

More WordPerfect Metadata. Some metadata easily can be accessed in WordPerfect while other types become visible by opening the document in a lower version of the software or in a binary file editor.

Other examples of metadata in WordPerfect include:

- Comments — these also can include information you don't want others to see.
- Company or organization name.
- Document revisions.
- File properties — summary tab.
- Hidden text.
- Initials (stored in "Tools," "Options," then "General").

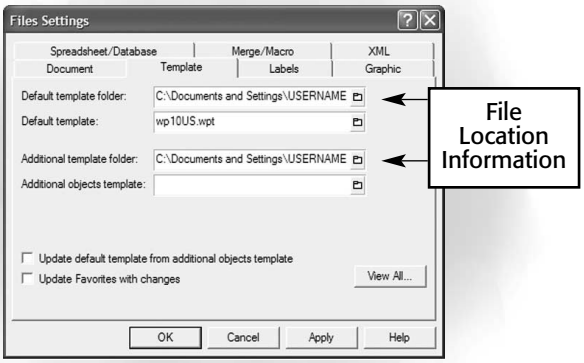


Figure 3. Username can be found in File Settings.

Once removed, the information is only removed in newly created documents, not existing ones.

- Non-visible portions of embedded OLE objects.
- Other file properties and summary information.
- Revisions and annotations.
- The name of your computer.
- Username (stored in "Tools," "Options," then "General"). Again, once removed, the information only is removed from newly created documents, not existing ones (see Figure 3).

To my knowledge, there is no tool that automatically will remove all metadata stored in WordPerfect documents. In the past, Corel has made available a utility on its File Transfer Protocol site that removes Undo/Redo History along with unused styles from documents. However, all other information must be removed manually.

Microsoft Word Metadata

Payne Consulting Group first became acutely aware of metadata when one of our lawyers sent a contract to us originally created for another technology firm. They used the previous client's document as a boilerplate form to create a new contract for us, replacing the other client's information with ours.

Most attorneys create contracts in this manner, and it makes sense. Why start from scratch each time when the intellectual property required to create such an agreement already exists? In this case, however, a corrupted document caused information to be compromised due to no fault of the firm or practicing attorney.

Word includes traditional metadata (not visible to the naked eye without clicking a button or two), extenuating metadata meant to include tracked changes not visible, versions, file properties and other blocks of information. The following is a list of some of the more common metadata in Word documents:

- Attached template
- Author
- Category, keywords and comments
- Company name
- Custom Properties such as client and matter or docID
- Date the file was created, last modified and printed
- Edit time
- Embedded objects
- Full name and file path where the document and template reside
- Graphics and more
- Hidden text
- Last 10 authors
- Manager
- Residual tracked changes
- Routing slip
- Subject
- Title
- Versions.

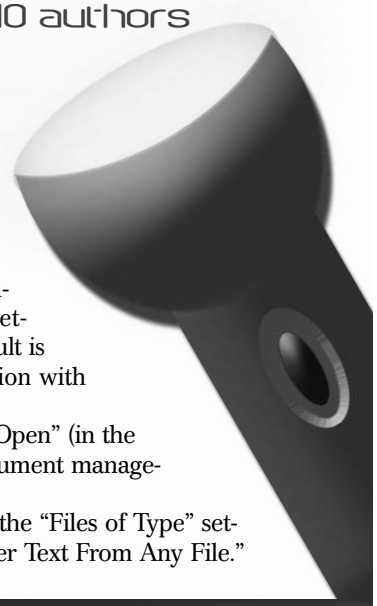
The following scenarios allow you to view examples of various forms of metadata saved in Word documents.

Viewable Edit Time. You draft a contract for a client. The client invoice is generated for the work and sent to the client. The client opens the contract sent electronically and does the following:

1. From the "Insert" menu, choose "Field."
2. Under "Field Name," select "EditTime."
3. Click "OK." The amount of time spent editing the document is inserted at the active cursor location.

While most clients understand not all billable time is spent constructing the document, they might question why they are charged X number of hours when the metadata about edit time specifies something very different.

Word includes traditional metadata (not visible to the naked eye without clicking a button or two) such as author, hidden text, company name, last 10 authors and much more.



Last 10 Authors. Before Microsoft Office XP, it was very easy to view the last 10 authors, full name and file path where the document was stored locally or on a network. Even with Office XP, the default is still to save and store this information with the file.

1. From the "File" menu, choose "Open" (in the native application if using a document management system).
2. Navigate to the file, but change the "Files of Type" setting in the dialog box to "Recover Text From Any File."

If you use Word 97, this feature might not have been installed in a typical installation, but can be installed with supplemental conversion file filters. Word 2000 and above install the file type by default.

3. Select and open a Word document created a while back.
4. Scroll toward the bottom of the document. You might see a list of names and file locations of people who previously worked on the document.

In Figure 4, you can see some of the metadata uncovered that was previously invisible before using this method. Of particular interest are names, Donna Payne and Karen Walker, along with file location information. Additional meta-

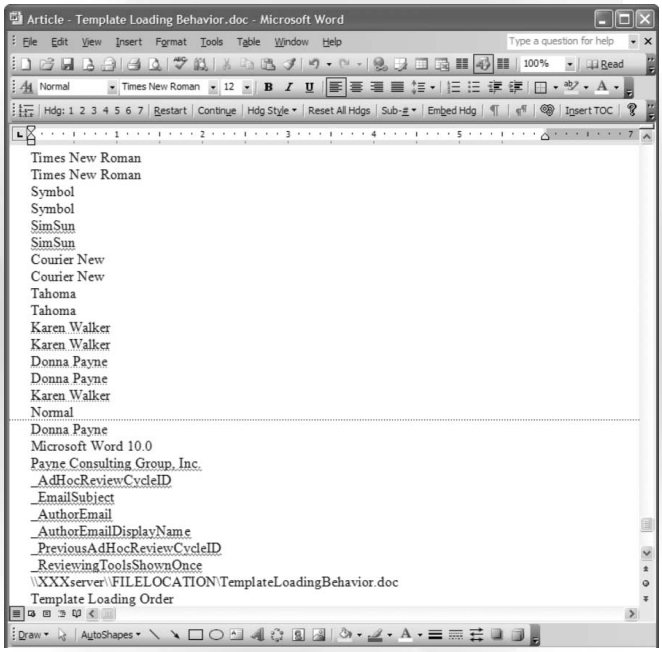


Figure 4. View last 10 authors.

data is visible regarding whether the file was e-mailed, including the sender's name, e-mail address and subject.

Embedded Objects. When an object is copied from a Web page or embedded into a document as an object, metadata associated with that object is attached. There was an instance where a person giving a presentation copied and used unauthorized graphics from a Web site. The problem? The owner of the reserved mark was in the audience, and soon after a cease and desist order was issued. Embedding spreadsheets introduces even more problems.

1. Open Excel and enter some text in several of the cells.
2. Switch to a different sheet in the workbook and type "This is very confidential."
3. Return to the initial sheet and select and copy the information.
4. Switch to Microsoft Word.
5. From the "Edit" menu, choose "Paste Special."
6. Select "Microsoft Office Excel Worksheet Object" and click "OK" (see Figure 5).
7. Double-click the inserted data from the Excel spreadsheet and notice the copied and pasted data is visible and the other sheets in the workbook also are available to view.

File Properties. As you work, information is added and saved to the document without your knowledge. The "File

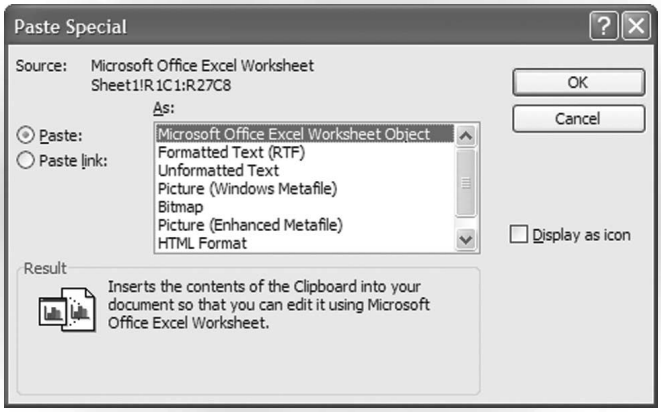


Figure 5. Copied and pasted data is visible in a new document.

Properties" dialog box is one place to look for some of the embedded information.

1. Create a new document.
2. At the top of the document type "This is very confidential and must be deleted before sending to client" (See Figure 6).
3. Save the file natively (not through the document management system).
4. Now delete the text from the document.
5. From the "File" menu, choose "Properties," then "Summary." The original typed text was saved and added to the document summary. Even though the information later was deleted from the actual file, it remains in the document properties.

Figure 6. File Properties.

File Versions. In Word, the concept of "Versions" means something different than what we have come to know in the legal community. Word allows you to save a version of the document with the same name inside the file. The document then can be rolled back to previous versions. The individual version history, along with who made the changes, are saved entirely within the same document.

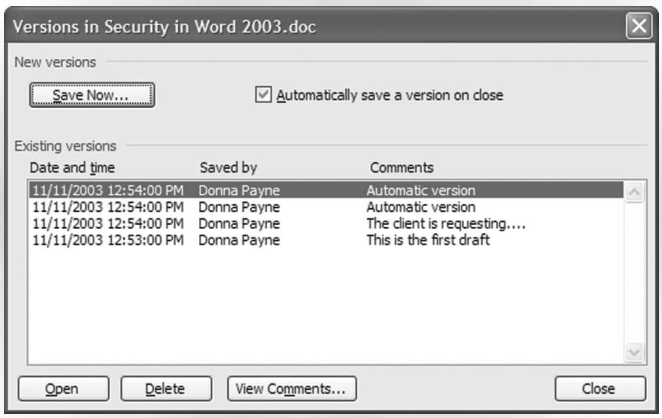


Figure 7. Microsoft Versions.

Not many law firms use Word's "Versions" feature, but someone with unscrupulous intent can turn on the feature and then forward the document to someone else. When the file is returned, a running history of every change made each time the document is saved travels with the document.

Controlling Versions. Note: Many firms remove the "Version" command from the "File" menu.

1. From the "File" menu, choose "Versions."
2. Check the option to "Automatically save a version on close." The first time you do this, you are prompted to add comments. After this, the feature kicks in automatically.
3. Save the file and close it.
4. Reopen the file and make changes. Save and close, then reopen the file again.
5. From the "File" menu, choose "Versions" and note each time the document was edited, saved and closed, a new version of the file was created (see Figure 7).

If your firm removed the "Versions" command from the "File" menu, make sure to check the Status bar of the Word window when you work on a document that came from out-

Figure 8. Double-click to display Versions.

side of the firm. If "Versions" is enabled, an icon with three diskettes will appear (see Figure 8). Double-click the icon to display the "Versions" dialog box.

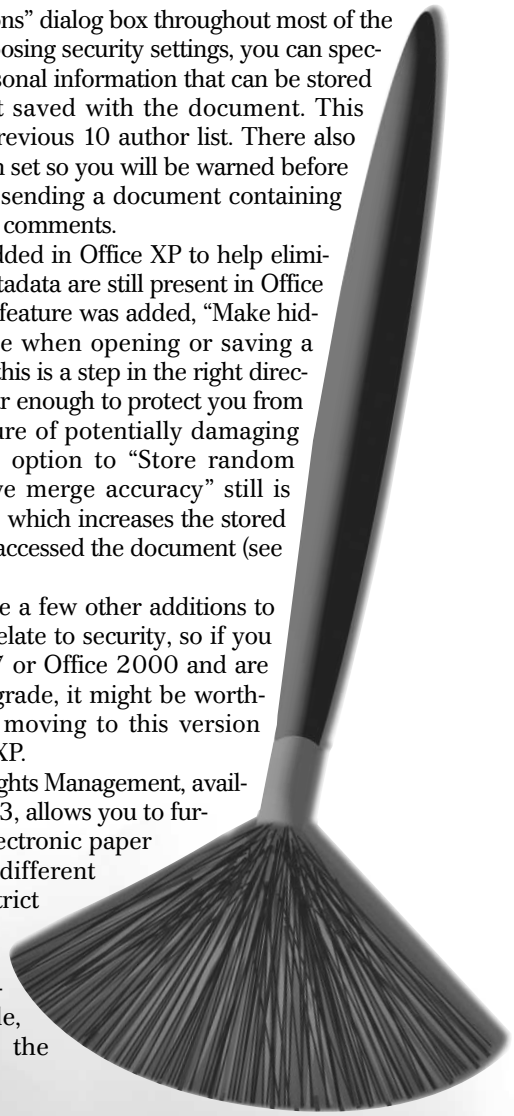
Word 2002 and 2003 Security Options to Minimize Metadata

Office XP introduced some help to fight accidental disclosure of confidential information. Security was further enhanced in Word/Office 2003. A new "Security" tab was added to the "Tools," then "Options" dialog box throughout most of the Office Suite. By choosing security settings, you can specify some of the personal information that can be stored as metadata is not saved with the document. This takes care of the previous 10 author list. There also is an option you can set so you will be warned before printing, saving or sending a document containing tracked changes or comments.

The options added in Office XP to help eliminate or identify metadata are still present in Office 2003. Also, a new feature was added, "Make hidden markup visible when opening or saving a document." While this is a step in the right direction, it doesn't go far enough to protect you from accidental disclosure of potentially damaging data. Further, the option to "Store random number to improve merge accuracy" still is checked by default, which increases the stored history of who has accessed the document (see Figure 9).

There are quite a few other additions to Office 2003 that relate to security, so if you are using Office 97 or Office 2000 and are considering an upgrade, it might be worthwhile to consider moving to this version rather than Office XP.

Information Rights Management, available for Office 2003, allows you to further restrict the electronic paper trail by assigning different permissions to restrict viewing, distribution and even printing of documents. For example, you can specify the



Security Tab Options in Word 2003

Option	Result
Remove Personal Information from File Properties on Save	Document specific; must be turned on for each document you work on; helps to remove some information from the "File Properties" dialog box, such as last 10 authors or author names when track changes or comments have been made to the document.
Warn before printing, saving or sending a file that contains tracked changes or comments	Displays a message box indicating the document has tracked changes or comments. This option only prompts when the document is open and you are working in Word. If you attempt to send a document that contains tracked changes through alternate-clicking and choosing "Send To: Mail Recipient," or if you send through your document management system or a third-party comparison program, the option is ineffective.
Store random numbers to improve merge accuracy	According to Microsoft help in Word 2003, this option "instructs Word to use randomly generated numbers to help keep track of related documents for comparing and merging." Although these numbers are hidden, they could potentially be used to demonstrate that two documents are related. If you choose not to store these numbers, the results of merged documents will be less than optimal. This isn't very reassuring; unless you use Word's built-in "compare and merge" feature, uncheck this option.
Make hidden markup visible when opening or saving	This feature displays tracked changes or comments in a document when it's opened.

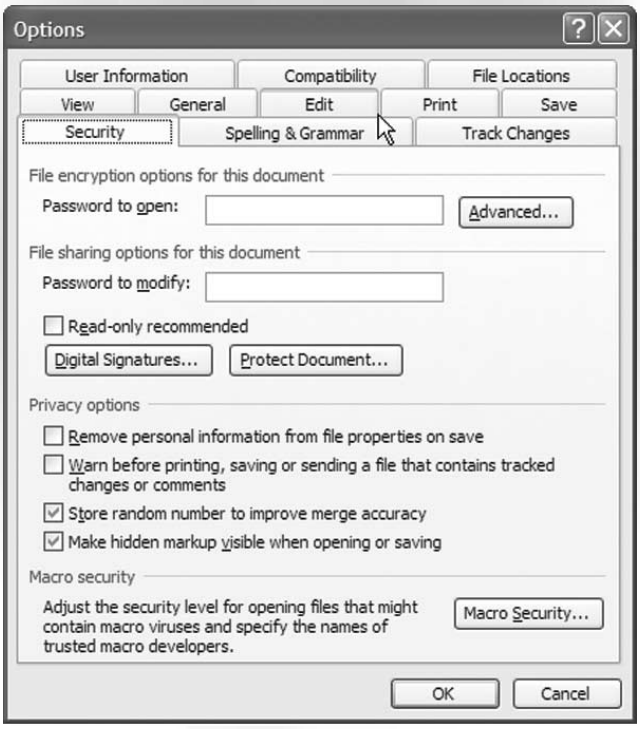


Figure 9. "Store random number to improve merge accuracy" option.

document only is to be printed a specific number of times to limit the amount of hard copies produced. Expiration dates can be set to have the file expire and not be opened after a time limit. Security changes and now security features in Office 2003 are worthy of an article all its own.

Other Microsoft Office Applications and Metadata

Metadata exists in other Office applications and most other software. Excel contains file properties similar to Word, along with special ways to format cells to be invisible, hide rows, columns and even worksheets. PowerPoint also has file properties, hidden slides, tracked changes, speaker notes and more. Microsoft Outlook includes settings to track all files sent and to then embed a history or electronic trail.

Several resources are available to help you understand how to minimize or eliminate metadata from Microsoft Office applications. Third-party tools for removing metadata from Microsoft documents are available, such as Metadata Assistant from Payne Consulting Group (the company I work for). Other companies with products to combat metadata are Workshare, Kraft Kennedy & Lesser, Esquire Innovations, and probably a few other vendors by now.

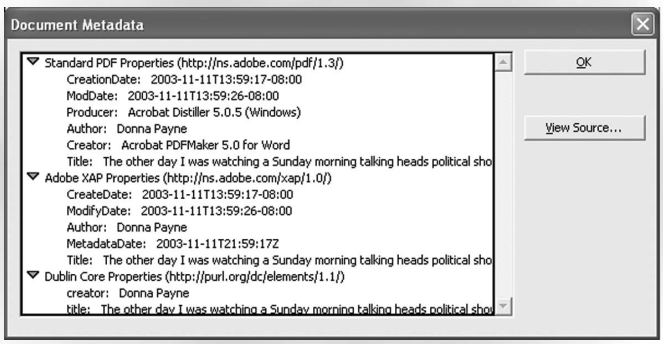


Figure 10. PDF document metadata.

For Knowledge Base articles on the subject, go to Microsoft.com and click "Support," then "Knowledge Base." Search for the keywords "How to Minimize Metadata" and select the appropriate version and application (Word, Excel, PowerPoint).

Metadata in Adobe PDF Files

Many firms create and send Portable Document Format files instead of Word documents to cut down on outside sources being able to view metadata. While this sounds like a good idea in theory, it's not always practical. It's common for documents that require substantial editing to float between attorney and client before finalization. And coming from the client side, I prefer all of my legal documents be in an editable format for potential future use.

While PDF files don't have the amount of metadata found in Word files, there still is substantial information gathered (see Figure 10).

Examples of PDF Metadata:

- Title
- Author (user ID)
- Document summary (see Figure 11)
- Keywords
- File location
- Comments and tracked changes, if they are contained in the original document.

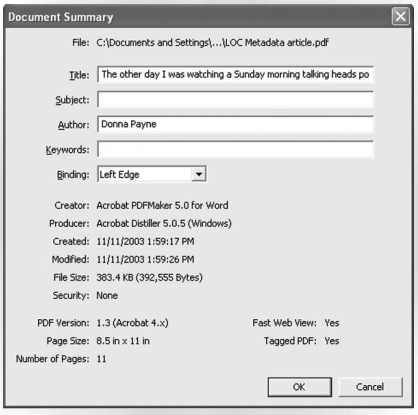


Figure 11. PDF Document Summary.

If you save files to PDF, it's important to clean out the original software program's metadata prior to converting to a PDF format so the information isn't automatically transferred.

To further secure the PDF file, open the PDF file and apply a password (under the "File" menu, chose "Document Security," "Security Options" then "Acrobat Standard Security"). You will want to apply a password to "Change Permissions and Passwords," and under "Permissions," don't allow the reader to change the document, copy or extract information, or add or change comments and form fields.

Conquering Metadata

As more documents are shared electronically, leaving unknown metadata in documents becomes a more pressing problem. To avoid embarrassing disclosures or malpractice know what you are sending outside the firm before clicking "Send." LOC

ABOUT THE AUTHOR

DONNA PAYNE is president and founder of Payne Consulting Group, a training and development company headquartered in Seattle. Payne has authored 10 books on Microsoft Office including the bestselling series: "Word for Law Firms." She is a member of Microsoft Legal Advisory Counsel, American Bar Association and the American Society of Journalists and Authors. Payne is a frequent speaker for conferences and product launches worldwide. Payne is a columnist for several legal and technical publications, and has been featured in syndicated articles on women in technology. You can contact Payne at: donnapayne@payneconsulting.com.