



# The New Metadata Rules

What a busy attorney won't take the time to tell you, and how it affects the legal IT department.

by DONNA PAYNE, PAYNE CONSULTING GROUP

**T**en years ago, I had my first lesson in metadata rules, and it's one I'll never forget. Our lawyer had created a product contract based on others that he and the firm had used for similar companies. He accepted the tracked changes, customized the version for us, and then e-mailed the document. When I opened the document, I discovered that all of the changes that he'd previously accepted were now back – and visible. The document had been corrupted and the tracked changes were again part of the document.

From that day forward, a large part of my life has been consumed with gathering as much information as possible about the subject of metadata.

Fast forward to today, and we're still breaking new ground. There are new opinions, rules and precedents being set for dealing with electronic documents and native production.

The problem is that while IT people have the technical know-how about metadata and how information is stored and accessed in electronic documents, attorneys have the knowledge about court rules and opinions. And, unfortunately, it's rare to find a firm who is communicating this information in both directions.

This article revisits the subject of metadata and examines some of the court rules, opinions and cases that have set legal precedent. It covers some of the additional efforts made by Microsoft in Office 2007 with respect to metadata, and it concludes with recommendations for keeping your metadata IQ up to date.

## Understanding Metadata

Metadata is formally defined as "data about data." Just think of it as information that is saved and stored with the file, which later can be used to locate documents based on the supplied criteria. For instance, you might need to search for any document created within a specified timeframe, say the last seven days. Your search engine would use file metadata to accomplish this. Perhaps you need to find any document created by a certain author or about a specific subject — again metadata. So metadata isn't necessarily a bad thing, because it has some useful purposes. It's only bad when it includes information you don't know is there that should remain confidential.

There is metadata that is added automatically either by the operating system, application or through other methods that occur without user intervention. Then there is metadata that is introduced by authors or reviewers of the document. This is where many organizations have gotten into hot water. Track changes and comments have led to many unintentional disclosures.

Take the United Nations document that contained residual tracked changes about an assassination and their conclusion on who was behind it, or SCO Group's lawsuit against DaimlerChrysler which originally identified Bank of America as the defendant instead of the automaker. I cannot tell you how many times I've used Google to search for random documents and found residual tracked changes embedded within the file. Most of the time, with the exception of when a file corrupts, the tracked changes and comments were left there unintentionally by the author or reviewers of the file. The more hands a file is

passed through, the higher the probability of information being left in the file unintentionally.

### Not the Same Metadata as Before

Microsoft Office 2007 goes further than any previous version of the Microsoft Office software in protecting from accidental disclosure because it collects less metadata and provides a methodology for removing certain types of metadata. It doesn't get all of it; for that you'll need a commercial metadata removal tool.

In Word 2007, there are some types of metadata that aren't processed any longer. For example, Word's Versions feature is a thing of the past, as are routing slips and Fast Saves options. Author information is minimized, and you can even set installation options to reduce the metadata footprint on your files.

For instance, you can set policies such as Do Not Track Document Editing Time, Disable Inclusion of Document Properties in PDF and XPS, and Enforce PDF Compliance with ISO 19005-1.

Within the application itself, you can turn off the capture of certain metadata types. For instance, in Word 2007, click the Office button, then Word Options. Select Trust Center, then click Trust Center Settings and select Privacy Options. Here you can specify document settings such as Warn Before Printing, Saving or Sending a File that Contains Tracked Changes or Comments (good), Store Random Number to Improve Combine Accuracy (turn off to eliminate this metadata from being added), and Make Hidden Markup Visible When Opening or Saving (good). The Remove Personal Information From File Properties on Save is a legacy setting. It's no longer necessary in Office 2007 because this is the default for all documents. Within Outlook, choose Tools, Trust Center, Attachment Handling to uncheck the option Add Properties to Attachments to Enable Reply with Changes.

### Using the New Document Inspector

Office 2007 includes the big sister of its legacy tool, Remove Hidden Data, but now it is built into the product. You can inspect the document to see if any metadata exists and then remove it. It doesn't show you the metadata found; its sole purpose is to identify whether or not specific metadata exists in the file and then offer the methodology of automatically removing it.

Unfortunately, the Document Inspector has a couple of drawbacks. First, it doesn't run automatically, so you have to remember to run it on each document. Commercial products that are available for purchase often integrate with document management systems and e-mail programs, so when you click Send, the file is cleaned, or you are at least reminded to clean it prior to its being sent. The other downside is that when you inspect a document, it lumps certain types of metadata together. For instance, Headers, Footers (including page numbers) and Watermarks are all handled under one grouping, so it's an all-or-nothing cleansing. The same is true of Comments and Tracked Changes when, in fact, you might want

## Checklist for Protecting Yourself and the Firm

If you're using an earlier version of Microsoft Office, run, don't walk, to **upgrade to Microsoft Office 2007**. It does a better job of minimizing metadata.

You should suggest or **implement a firm policy on metadata and data privacy**. Include a statement from management on the strong commitment. Also include a statement on the policy of handling inadvertent disclosure. If you're interested in seeing a draft of a policy from Payne Consulting Group, send an e-mail to [donna-payne@payneconsulting.com](mailto:donna-payne@payneconsulting.com).

If you outsource the word processing function or help desk, you still need to **ensure that your subcontractors are compliant** since they are acting on behalf of the firm and handling sensitive documents.

Don't touch documents that are part of a litigation or are reasonably expected to become part of a litigation. This could result in spoliation. It's best to **clean copies of documents and leave the original intact**.

**Purchase or develop a metadata removal tool**, and make sure it extracts the most metadata possible from your documents. It doesn't matter how many bells and whistles a product has, or how aesthetically pleasing it is, if it doesn't get the full measure of metadata, it can compromise you and give you a false sense of security.

**Stay current on bar opinions and court cases that cite metadata.** (See sidebar on page 11, "Legal Rules, Opinions and Precedents on Metadata.")

**Educate yourself and your firm.** There is no shortage of seminars on metadata and inadvertent production and disclosure. Sign up for one. Even if you know a lot about the subject, you can always learn something new. As you learn new things, weave this into learning for your users. Offer continuous and updated training, and don't assume everyone is following the policy just because one has been put into place.

**Check documents in the public domain.** You'd be surprised what you find when you search for your organization's name and set certain criteria in a search engine such as Google.

to keep one or the other independently when collaborating with co-counsel or a client.

Finally, not all of the metadata is removed, but it does get a large portion of it. Combine Inspect Document with the action of saving the cleaned document as a PDF file and you have a workable, barebones method for removing most of the embedded metadata.

According to Microsoft's security policies and settings in the Office 2007 system, you cannot individually disable the Inspector module for Comments, Revisions, Versions and Annotations, or the Inspector module for Document Properties and Personal Information. To disable a Document Inspector, type the CLSID for the inspector you want to disable (the inspector CLSID can be found at HKLM\Software\Microsoft\Office\12.0\Word\Document Inspectors and similarly in Excel and PowerPoint).

In summary, if you are using a third-party metadata removal tool, due to the automation and integration as well as choices for what is removed and retained, I think you'll still want to keep this solution in place — at least for now.

## When Redaction Fails

I have documents in my possession that were improperly redacted. One even includes secret Grand Jury testimony that, instead of being truly redacted, was highlighted. Then there are others from both the Justice Department and the Pentagon. This just shows that no matter how security-conscious an organization is, exposure can happen.

A few years back, I was giving a speech for a convention center full of judges, and the person speaking ahead of me spent a good deal of time talking about how to redact properly. He said, "Just select the text, make the background black, and the text set to match the background – then convert it to PDF." While this is possible, it's not advisable because when you select, copy and paste this information from the PDF file to Word or another text reader, the redacted text can become exposed. That's why it's important to use a tool for redacting your confidential documents the right way. I know of two products that do a good job with redaction. The first is Adobe Acrobat, which includes secure redaction. The other is a product from DocsCorp. I'm sure there are others, but I have personal experience with these, and they work.


You might ask, why can't I just use a marker and markup the text, then scan it? If you do this correctly and then check the result, you can definitely redact this way. It's always important to check the scanned file. I have seen scanned, redacted documents where I've been able to clearly read the text through the markup, because the scanner equipment was powerful enough to read the text through the markup. Instead, if you want to resort to the low-tech method, just put a piece of correction tape over the text, on both sides of the paper, and then scan. This should do the trick. Once again, make sure you check the scanned copy to ensure that nothing is visible that should not be seen.

Before you think making a PDF solves all metadata problems, think again. At Payne, we recommend that you clean the file in the native application first, and then convert it to PDF

file format. This adds an extra layer of protection to remove embedded information that might be carried over into the PDF file. For instance, I once saw a Department of Defense document that included an Army Colonel's direct e-mail address and confidential information because the document had been attached to an e-mail message and sent, thus embedding the sender information along with the file.

If you're going to use PDF, add protection to prohibit copying or modification of text. I recommend taking some time to read and learn about the different levels of protection in PDF files to make sure you're doing everything you can to protect yourself.

## Worth the Time to Discuss

The subject of metadata is finally something that we can share with attorneys because it's an IT matter that has relevance in the courts. Take time to discuss this issue with them, and keep them involved in new discoveries that you find. 

This article was first published in ILTA's October, 2008 white paper titled "Microsoft — The Promise of New Technology" and is reprinted here with permission. For more information about ILTA, visit their website at [www.iltanet.org](http://www.iltanet.org).

## Legal Rules, Opinions and Precedents on Metadata

The technical aspect of metadata is pretty easy to grasp. The problem for many of us in IT is keeping in touch with the ever-changing rules governing the viewing, maintenance, production and discovery of electronic documents and their associated metadata. Add to this the conflicting positions by the American Bar Association and some of the state bar associations, and keeping up with the rules becomes almost a full-time job.

The American Bar Association Formal Opinion 05-437 states that you have a duty to notify the sender of the inadvertent production. At least 19 jurisdictions follow this rule. Other states take a different stance, which can be anywhere from lenient to hard-line. Still other states have remained silent (at least at this time) on how metadata and inadvertent disclosure should be handled. The following are some of the existing rules and positions on the subject.

**Colorado Bar Association Ethics Opinion 119:** A lawyer should generally be allowed to look at metadata, unless the attorney knows or has reason to believe that the document contains privileged information.

**Arizona Bar Association Ethics Opinion 07-03:** While lawyers must use reasonable care to “scrub” metadata from outgoing documents, a recipient must generally avoid looking for inadvertently included metadata and must also notify the sender of the inclusion of any inadvertent metadata.

**D.C. Bar Ethics Opinion 341:** “A receiving lawyer is prohibited from reviewing metadata sent by an adversary only where he has actual knowledge that the metadata was inadvertently sent. In such instances, the receiving lawyer should not review the metadata before consulting with the sending lawyer to determine whether the metadata includes work product of the sending lawyer or confidences or secrets of the sending lawyer’s client.”

**Alabama State Bar Ethics Opinion 2007-02:** “(a) A lawyer shall not reveal information relating to representation of a client unless the client consents after consultation, except for disclosures that are impliedly authorized in order to carry out the representation, and except as stated in paragraph (b).” And, there is a duty under Alabama Rules of Professional Conduct, Rule 1.6 to use reasonable care when transmitting electronic documents to prevent the disclosure of metadata containing client confidences or secrets. Absent express authorization from a court, it is ethically impermissible for an attorney to mine metadata from an electronic document he inadvertently or improperly receives from another party.

**Florida Bar Ethics Opinion 06-2:** This opinion requires that “a lawyer who is sending an electronic document should take care to ensure the confidentiality of all information contained in the document, including metadata. A lawyer receiving an electronic document should not try to obtain information from metadata that the lawyer knows or should know is not

intended for the receiving lawyer. A lawyer who inadvertently receives information via metadata in an electronic document should notify the sender.”

**New York State Bar Association Ethics Opinion 749:** Requires metadata removal from electronic files. The New York State Bar Association concluded that the use of computer technology to access client confidences and secrets revealed in metadata constitutes “an impermissible intrusion on the attorney-client relationship in violation of the Code.”

**Pennsylvania Bar Association Ethics Opinion 2007-500:** “It would be difficult to establish a rule applicable in all circumstances and that consequently the final determination of how to address the inadvertent disclosure of metadata should be left to the individual attorney and his or her analysis of the applicable facts.”

**Texas Rules of Civil Procedure 196.4:** The requesting party must specifically request electronic data and the form in which it should be produced, and the responding party must produce it in that form or state an objection.

Regardless of where you are located, it’s important to stay on top of the rules and opinions because they have a way of creeping into other jurisdictions in some way, shape or form.

You should also be aware of Amendments to the Federal Rules of Civil Procedure:

Rule 26(a)(1)(B) — Requires initial disclosure of sources of discoverable information.

Rule 26(b)(2)(B) — Requires that parties identify sources that are “not reasonably accessible because of undue burden or cost.”

Rule 26(b)(5)(B) — Requires that parties sequester privileged/non-discoverable information; inadvertent production is not a waiver of privilege.

Rule 26(f) — Requires that parties meet early in the discovery process to discuss “any issues relating to preserving discoverable information.”

Rule 34(a) — Describes the scope of production information; e-mail messages are discoverable.

Rule 34(b) — Describes the procedure by which to request information; the “request may specify the form or forms in which electronically stored information is to be produced.”

To stay on top of rules and opinions, consult such sites as:

[www.legalethics.com](http://www.legalethics.com)  
[www.payneconsulting.com](http://www.payneconsulting.com)  
[www.abanet.org/cpr](http://www.abanet.org/cpr)  
[www.thesedonaconference.org](http://www.thesedonaconference.org)  
[www.microlaw.com](http://www.microlaw.com)