

METADATA:

What It Is & Why You Should Care

Susan J. Silvernail, Esq.  
Marsh, Rickard & Bryan, P.C.  
ALAJ Summer Seminar  
August 11, 2007

## **“DATA ABOUT DATA”**

Metadata is commonly defined as “data about data”. But unless you are already a little techno-savvy, that definition probably is not terribly helpful to you. Try this define : metadata is data that is attached to a computer file that describes the file. Think of it as extra information that is hidden in a document, information that is automatically created and embedded in a computer file.

So, every time a document is created or amended in Microsoft Word, Excel or PowerPoint, data tracking the author, document changes, editing time, and other document properties are added to the document. <sup>1</sup> On its website, Microsoft indicates that the following metadata may be stored in documents created in all versions of Word, Excel and PowerPoint:

- Track changes: Inserted or deleted text you thought was gone
- Speaker notes
- Hidden cells
- Comments
- Your name
- Your initials

---

<sup>1</sup> There are many different types of metadata, including metadata about websites, metadata about video files, and metadata about electronic documents. For ease of discussion, the focus of this paper is document metadata, which describes document attributes such as the title, author, content, location, and date of creation. E-mail carries metadata, identifying the path it took from origin to destination, along with times, Internet Protocol addresses and the like.

- Your email address
- Your company or organizations's name
- The name of your computer
- The name of the network server or hard disk on which you saved the document
- Other file properties and summary information
- The names of previous document authors
- Document revisions
- Document versions
- Template information
- Hidden text
- Macros
- Hyperlinks
- Routing information
- Nonvisible portions of embedded Object Linking and Embedding (OLE) objects

Similar, although less, metadata exists within Corel WordPerfect files, and metadata security issues affect the documents created in most other software programs, as well.

Some metadata can easily be viewed within the program that has created a file. In most circumstances, hidden metadata can only be seen

with special software. However, hidden metadata can become visible accidentally, for example, when a corrupted file is opened or when WordPerfect opens and improperly converts to a Word file.

The metadata remains present, but hidden, until you remove it or someone else extracts it. And therein lies the danger: the wrong eyes may see this metadata. If you have e-mailed a Word or WordPerfect document to either a client or opposing counsel, chances are high that you shared more information than you intended.

### **MIND YOUR METADATA**

Here are some high-profile examples where someone sending an electronic document to someone else forgot that hidden metadata exists:

*October 2000:* The Wall Street Journal reports that a candidate running for the U.S. Senate began receiving anonymous emails containing messages written in MS Word criticizing and attacking the candidate. A savvy aide looked at the document properties and discovered they were authored by the chief-of-staff of the opposing party.

*February 2003:* A dossier on Iraq's security and intelligence organizations, cited by Colin Powell and published by 10 Downing Street, is discovered to have been plagiarized from a U.S. researcher on Iraq. Since the dossier was published on their website in MS Word format, researchers also discovered the four people in the British government who edited the document. They were subsequently called to Parliament for a hearing.

*March 2004:* SCO Group, seller of UNIX and Linux, sent out a warning letter to 1,500 of the world's largest companies threatening legal liability for using Linux if they failed to obtain a

license from the Utah-based company. After filing suit against Daimler-Chrysler, metadata in a MS Word version of the suit revealed that the SCO's attorneys had spent a good deal of time aiming the suit at Bank of America instead.<sup>2</sup>

*May 2005:* Derrick Max, the head of two supposedly independent, nonpartisan groups who support overhauling Social Security, e-mailed his testimony on Social Security to the Senate but forgot to turn off "track changes". Turns out the associate commissioner of the Social Security Administration, who was on loan, working out of the White House, edited Max's "independent" testimony.

*October 2005:* United Nations issued report on Syria's suspected involvement in the assassination of Lebanon's former prime minister, Rafik Hariri. Recipients of a version of the report were able to track the editing changes, which included the deletion of names of officials allegedly involved in the plot, including the Syrian president's brother and brother-in-law<sup>3</sup>

*December 2005:* President Bush delivers speech at the U.S. Naval Academy outlining a new "National Strategy for Victory in Iraq", which was posted on the White House website. The New York Times, using Adobe System's Acrobat software that shows the document was created by "feaver-p" reveals that the speech was largely written by a Duke University political scientist, Peter D. Feaver.<sup>4</sup>

## **METADATA IS YOUR FRIEND**

The risk of accidentally revealing sensitive or confidential information

---

<sup>2</sup> See CS Reach, "Lemon Juice, Cornstarch, and Microsoft: Invisible Ink and Your Documents", American Bar Association Legal Technology Resource Center, 2007

<sup>3</sup> See T Zeller, "Beware Your Trail of Digital Fingerprints", New York Times, November 7, 2005.

<sup>4</sup> See S Shane, "Bush's Speech on Iraq War Echoes Voice of an Analyst", New York Times, December 4, 2005.

through metadata certainly makes for great headlines. It is easy to forget that metadata serves a purpose and is very useful. Document metadata categorizes information to make it easier to save and retrieve documents.<sup>5</sup> Document management systems make extensive use of metadata. Within Microsoft Word, the ability to view Comments and Suggested Changes can be very helpful when working with others to create a document. Yet, the collaboration features result in significant amounts of metadata being included in documents. For example, changes that are not accepted still remain with the document, even though they appear to be invisible. These changes can easily be displayed by turning on the "Show Markup View".<sup>6</sup> As one writer put it: "The problem is not that metadata is added to documents. The problem is that it cannot be easily removed from documents."<sup>7</sup> Now that the Alabama State Bar has spoken to the issue of metadata, however, there is little choice but to actively guard against the exposure—accidental or deliberate—of metadata.

## **EO-2007-02**

This summer, the Alabama State Bar Office of the General Counsel

---

<sup>5</sup> [Http://www.metadatarisk.org/document\\_security/doc\\_security\\_overview.htm](http://www.metadatarisk.org/document_security/doc_security_overview.htm)

<sup>6</sup> Id.

<sup>7</sup> M. Silver, "Microsoft Office Metadata: What You Don't See Can Hurt You", <http://articles.techrepublic.com>, March 4, 2003.

issued a formal opinion (2007-02) on the disclosure and mining of metadata. In doing so, Alabama became one of only a handful of states to deal head-on with the metadata issue but also, embraced a view contrary to that of the American Bar Association.

The first question addressed by Opinion Number 2007-02 is this: "Does an attorney have an affirmative duty to take reasonable precautions to ensure that confidential metadata is properly protected from inadvertent or inappropriate production via an electronic document before it is transmitted"? And the short answer is: **"Lawyers have a duty under Rule 1.6<sup>8</sup> to use reasonable care when transmitting electronic documents to prevent the disclosure of metadata containing client confidences or secrets"**.

The Opinion begins by defining metadata for these purposes as "data hidden in documents that is generated during the creation of these documents". Then, the Opinion acknowledges the risks attendant to the disclosure of metadata i.e., "the disclosure of client confidences and secrets, litigation strategy, editorial comments, legal issues raised by the client, and other confidential information". The Opinion provides two examples where client confidences or secrets could be at risk. The first is where a motion is

---

<sup>8</sup> See Rule 1.6 Alabama Rules of Professional Conduct, which provides that: "(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)."

written by several of a firm's attorneys using the collaboration features found in Word and WordPerfect and the motion is then electronically transmitted to opposing counsel. As the Opinion states, "[i]f you failed to 'scrub' or remove the hidden metadata prior to transmission, the opposing party could mine the document's metadata and discover which attorneys reviewed the motion, the critiques about the viability or strength of certain arguments, and the subsequent revisions made to the document". The second example involves the use of templates. Attorneys commonly base a document or a filing on a previous document. The Opinion points out that if the document is later electronically transmitted to the opposing party, the opposing party could discover the original client's name and information. The Opinion states that the "disclosure of client identity and information could constitute a violation of Rule 1.6, Alabama Rules of Professional Conduct".

With this ethical duty to exercise reasonable care when transmitting electronic documents in place, the issue then becomes what constitutes "reasonable care". The Opinion instructs as follows:

Factors in determining whether reasonable care was exercised may include steps taken by the attorney to prevent the disclosure of metadata, the nature and scope of the metadata revealed, the subject matter of the document, and the intended recipient. For example, an attorney would need to exercise greater care in submitting an electronic document to an opposing party than he or she would if e-filing a pleading with the court.

There is simply a much higher likelihood that an adverse party would attempt to mine metadata, than a neutral and detached court.

The second question addressed by Opinion 2007-02 is far more controversial: "Is it unethical for an attorney to mine metadata from an electronic document he or she receives from another party"? The short answer given by Alabama State Bar Office of the General Counsel is this:

**"Absent express authorization from a court, it is ethically impermissible for an attorney to mine metadata from an electronic document he or she inadvertently or improperly receives from another party".**

The Opinion defines "mining" the document as "the act of deliberately seeking out and viewing metadata embedded in a document". Then, the Opinion concludes that the unauthorized mining of metadata by an attorney to uncover confidential information would be a violation of Rule 8.4, Alabama Rules of Professional Conduct.<sup>9</sup>

This decision has the effect of making the burden to protect the confidences of a client on the receiving lawyer equal to that of the sending

---

<sup>9</sup> See Rule 8.4 Alabama Rules of Professional Conduct, which provides that it is misconduct for an attorney to, among other things: "(a) violate or attempt to violate the Rules of Professional Conduct, knowingly assist or induce another to do so, or do so through the acts of another; (b) commit a criminal act that reflects adversely on the lawyer's honesty, trustworthiness or fitness as a lawyer in other respects; (c) engage in conduct involving dishonesty, fraud, deceit or misrepresentation; (d) engage in conduct that is prejudicial to the administration of justice;"

lawyer. In reaching this conclusion, the Alabama State Bar Office of General Counsel expressly relied upon N.Y. State Bar Opinion 749 of the New York State Bar. In Formal Opinion 749, the New York State Bar wrote that “in light of the strong public policy in favor of preserving confidentiality as the foundation of the lawyer-client relationship, use of technology to surreptitiously obtain information that may be protected by the attorney-client privilege, the work product doctrine or that may otherwise constitute a ‘secret’ of another lawyer’s client would violate the letter and spirit of these Disciplinary Rules”. On that basis, Alabama State Bar Opinion 2007-02 states that “[t]he mining of metadata constitutes a knowing and deliberate attempt by the recipient attorney to acquire confidential and privileged information in order to obtain an unfair advantage against an opposing party”.

The Opinion carves out an exception for the mining of metadata that involves electronic discovery. The Opinion recognizes recent court decisions that indicate “that parties may be sanctioned for failing to provide metadata along with electronic discovery submissions”.<sup>10</sup> The Opinion provides an example in which metadata would be relevant and material to the underlying issues of a lawsuit: “the mining of an email may be vital in determining the

---

<sup>10</sup> See e.g., *Williams v. Sprint/United Management Company*, 230 F.R.D. 640, 652 (D. Kan 2005), where terminated employees brought a class action and sought Excel spreadsheets with all metadata intact, including embedded formulae. The court held that metadata ordinarily visible to users of Excel spreadsheets “should presumptively be treated as part of the ‘document’ and should thus be discoverable”.

original author, who all received a copy of the email, and when the email was viewed by the recipient". The Opinion wraps up by cautioning attorneys to seek direction from the court as to the production of metadata during discovery.<sup>11</sup>

In interpreting the American Bar Association's Model Rules of Professional Conduct, the ABA's Standing Committee on Ethics and Professional Responsibility declined to place a like-kind burden on lawyers receiving electronic information from opposing counsel. In November 2006, the ABA Ethics Committee issued Formal Ethics Opinion 06-440, which interprets Model Rule 4.4.<sup>12</sup> The Opinion concludes that lawyers who receive electronic documents are free to look for and use information hidden in metadata. The ABA Ethics Committee states that the rule:

"requires only that a lawyer who receives a document relating to the representation of the lawyer's client and who knows or reasonably should know that the document was inadvertently sent shall promptly notify the sender. The Rule does not require refraining from reviewing the materials or abiding by instructions of the sender."

---

<sup>11</sup> Be aware that if a document is subject to discovery, the document should not be "scrubbed" or have its metadata removed. Given the 2006 e-discovery amendments to the Federal Rules of Civil Procedure, prudent parties will reach an agreement on the format of production before scrubbing metadata. See E. Libby "What Lurks Within: Hidden Metadata in Electronic Documents Can win or Lose Your Case", ABA Journal, April 2007.

<sup>12</sup> Maryland's State Bar Association's Committee on Ethics issued an opinion, Opinion 2007-09, concluding, like the ABA opinion, that it is not an ethics violation to look at metadata received from opposing counsel.

Consistent with Alabama Ethics Opinion 2007-02, however, the ABA Ethics Committee did recommend that lawyers take steps to guard against the disclosure of metadata, including “scrubbing” metadata from document prior to disclosure and entering into “clawback” or non-waiver agreements.<sup>13</sup>

### **WHAT TO DO ABOUT IT**

As indicated above, being aware of metadata is a good place to start; but taking steps to reduce or eliminate the risk of document metadata is paramount. There are several options to reduce or eliminate metadata from your documents. The drawback is that there is no central place to manage the different settings. Users must go to several different places within the application to remove different types of metadata.

Word, PowerPoint and Excel users can turn off “Fast Saves” under the Save tab in the Options menu to ensure that deleted data is really deleted. The “Fast Saves” feature lets a computer more quickly save a file by not removing deleted text from it. Older versions of Microsoft Office products turn on “Fast Saves” by default. When computers were slower, it was a helpful feature but today, users are not likely to notice any difference with this feature turned off.

---

<sup>13</sup> “Clawback” or non-waiver agreements, under which inadvertently produced material is returned without waiver, are viewed with favor by the 2006 e-discovery amendments to the Federal Rules of Civil Procedure. See the Committee Notes to FRCP Rule 26(b)(5)(B) and 26(f)(4).

When using features such as tracked changes, document versions or comments, make sure to delete the information being kept within the document with these features.

Under the Security Tab in the Options menu, Microsoft provides the ability to remove personal information from a file upon Save and to warn users before printing, saving, or sending a file that it contains tracked changes or comments. But, Word only warns users sending documents from within Word; it does not warn if documents are attached from within Outlook or other mail programs.

Excel and PowerPoint can remove personal information but they do not warn about comments in either program or tracked changes in Excel.

Some technology writers have recommended that Word documents not be sent outside the immediate work environment at all and that a document conversion process be required as part of a "publication" activity.<sup>14</sup> One common method is to enforce conversion to PDF before sending externally. Converting files to PDF format with Adobe Acrobat or other PDF creators will strip out most metadata. Sending documents in a format that prevents a document from being edited can actually be helpful.

Converting a document to PDF format will strip out metadata from the original document; but PDF files can contain their own metadata. This is

---

<sup>14</sup> Id. See also "Beware the Dangers of Metadata", LAWPRO Magazine, June 2004.

usually basic information like the author, date of creation and file location. Select File, then Document Properties to view the summary metadata information within a PDF file. Using the Security Options settings, it is possible to further restrict how the document can be accessed, used, copied and printed.

Microsoft has released an add-in called the "Remove Hidden Data Add-in". It will permanently remove hidden and collaboration data from Word, Excel and PowerPoint files in Office XP and Office 2003 on a PC. With this add-in installed, you get a new Remove Hidden Data option on the File menu. This allows you to save a copy of the original document with all personal information removed. Since the original information is irrevocably removed, Microsoft recommends running the add-in just before publishing or distributing the completed document.<sup>15</sup> Office 2007 users do not need the add-in because the capability to find and remove personal information is built into all Office 2007 programs.<sup>16</sup>

To be extra safe, several third party software developers have products to help remove metadata from the Microsoft suite of products:

- Esquire Innovations: iScrub

---

<sup>15</sup> <http://office.microsoft.com/en-us/help/HA011400341033.aspx>

<sup>16</sup> E Bott, "What's Hidden in Your Word Documents?", <http://ww.edbott.com/weblog/?p=1693>, June 11, 2007. Similarly, Adobe Acrobat 8.0 has new functionality to help users remove metadata from documents.

- Kraft Kennedy & Lesser: ezClean
- Payne Consulting Group: Metadata Assistant
- SoftWise: Out-of-sight
- Workshare: Workshare Protect

There is no software program for easily and automatically removing metadata from WordPerfect documents. WordPerfect has a feature called Undo/Redo History that allows you to view what was cut, copied, and even deleted in a document. Click on the Option button, and then uncheck "Save Undo/Redo Items with Document" to turn it off. However, this does not remove all metadata.

Another solution is to use WordPad, a stripped-down word processor in Windows, or save the file in Rich Text Format (RTF).

As the legal community becomes more aware of metadata and the damage inadvertent disclosure can cause, the need to minimize the risks to ourselves and our clients becomes evident. Here are the steps most often recommended to ensure that the documents you send or share with others remain secure and confidential<sup>17</sup>:

- 1) Educate your staff about metadata concerns;<sup>18</sup>

---

<sup>17</sup> See R Farrar & S McClellan, "Metadata Management in Microsoft Office: How Firms Can Protect Themselves against Unintentional Disclosure and Misuse of Metadata" ABA Technology eReport, May 2006; K Rutsky, "Metadata: What You Don't Know About Your Documents Can Hurt You", ABA Journal, November 2005.

<sup>18</sup> In this effort, a great on-line resource can be found at [www.metadatarisk.org](http://www.metadatarisk.org). The website is sponsored by Workshare.

- 2) Establish a firm-wide policy on how to share or distribute documents;
- 3) Install a firm-wide metadata removal application;
- 4) Cleanse metadata *before* converting to PDF;
- 5) Distribute final published documents in a secure, metadata-cleansed PDF format; and
- 5) Consider sending documents in zip format with a zip password.

### **CONCLUSION**

In conclusion, metadata in electronic documents can pose serious risks if left unmanaged or ignored. It is important to educate ourselves about metadata and develop metadata control strategies. This is especially true in the face of the recently issued Ethics Opinion from the Alabama State Bar (2007-02), which states that “[l]awyers have a duty under Rule 1.6 to use reasonable care when transmitting electronic documents to prevent the disclosure of metadata containing client confidences or secrets”.