

THE RECORDER

Shielding Metadata in Discovery

Randall Farrar

The Recorder
March 07, 2011



Randall Farrar, Esquire Innovations
Image: courtesy photo

Those of us in the legal profession are familiar with the concept of "discovery," and the requirements set by the courts for complying with discovery demands. We also know that we are only required to provide the documents and data information set out in the discovery demand. Unfortunately, in providing electronic documents, you may "discover" that you are inadvertently disclosing more information than you realize.

METADATA

Microsoft Word is the most widely used document production application on the market today. Whenever a document is created, opened, or saved in Word, the document stores information known as metadata, which the user had no intention of including or disclosing. In some cases, the disclosure of document metadata may have potential for harm or embarrassment. Attorneys and law firms must take responsibility to understand and manage document metadata risks and potential "adverse disclosure."

The document metadata contained in a Word document other than the intended text doesn't necessarily create risk of adverse disclosure. In fact, some document metadata is necessary for the formatting or automation of documents. Tracked Changes, which contains metadata, is useful when collaborating with cooperator attorneys, but should not be shared with adversary attorneys.

Word metadata can be separated into two categories. The first is called hard metadata, the document metadata that is readily accessible through the user interface of Word and that others and I have been writing about for years. The second category is called subtle metadata, the imperceptible stuff, the document metadata that may disclose origination and originator, but only to the trained "geeky" eye.

HARD METADATA

Hard Metadata is that document metadata that many are familiar with, and that all metadata management applications address.

Author metadata. For the most part, author information in Word can be controlled. Word has five areas that collect author information: 1) user name; 2) user initials; 3) document author; 4) manager; and 5) last author.

Windows system dates. Whenever a file is copied or moved from a local drive to the network, the system dates will change and will be different from Word's document dates.

Built-in document properties. Word's built-in document properties contain information about a document that describes or identifies it in some way. Built-in document properties are stored in the document and are not a part of the author's content. They include information such as title, author name, subject, and dates that identify the document's subject matter or contents. Unlike custom document properties, built-in document properties cannot be deleted and in some cases (e.g., print date) cannot be changed.

Built-in document properties can be valuable for document identification for later use. For instance, built-in document properties can be used to search for specific documents based on their properties.

Print Date. The print date never changes in a Word document. This displays the last time the document was printed and does not change (until it is printed again), and cannot be changed. Duplicating and revising only perpetuates this property. A print date can be older than the created date.

Tracked changes. There are times when a user may not be aware that changes are being tracked (author, date, time, etc.) in Word documents. To be certain they are not, I recommend a two-pronged approach — accept all changes and use Word's warning system options (found in Word 2010 under the trust center settings): "Warn before printing, saving, or sending a file that contains tracked changes or comments," and "make hidden markup visible when opening or saving."

External document links. Word provides field codes that can be linked to other documents or through which data from one document can be inserted to and updated in another document. These links show path information and could lead to Information Disclosure.

Hidden text. Hidden text is text that is formatted as hidden, or made invisible by changing the font color or size.

SUBTLE METADATA

As I mentioned earlier, subtle metadata can disclose origination and even originator, but only to the trained person who is considered a Word expert. What amazes me about subtle metadata is that it is

not talked about or addressed in many metadata management applications. Even Microsoft Office's built-in metadata application, Document Inspector, doesn't address it.

Subtle metadata is faint and inconspicuous. At first, it may appear rather benign, but it's like a road caution sign, warning of upcoming danger ahead. I want to focus on two questions in regards to subtle metadata: What are the subtle metadata elements in a Word document? And what best practices approach will mitigate disclosure of subtle metadata?

Word has five areas considered subtle metadata:

1) Custom Styles; 2) Bookmarks; 3) Document Variables; 4) Template Names; and 5) Custom Properties.

Custom Styles. Word provides an array of built-in styles, Quick Style sets, and themes for formatting documents. When desired formatting choices are not available, the users can create custom styles to suit their needs. This is where disclosure is an issue. I've seen firms and attorneys create custom styles with very specific "firm" names. In addition to the firm name in the style name, I've even seen user names or initials in the style names as well. The styles travel with a document, which means specific naming conventions may disclose origination and originator.

Keep it simple. Try to use built-in style names instead of creating a custom style name. Built-in styles can be customized to suit formatting needs. Use a generic naming convention when naming custom styles. For example, instead of naming a style "SMH Attorney Signature," it could be more generic, e.g., "Atty Sig." Do not use the firm name or user initials in custom style names.

When documents are leaving the electronic walls of a firm, set the firm's metadata management application to normalize (rename) all custom style names to generic names such as Style1, Style2, Style3, etc.

Bookmarks. A Word bookmark identifies a location or a selection of text that is named to allow quick access for future reference. For example, a bookmark might be used to identify text in a large document to be revised later. Instead of scrolling through the document to locate the text, the user can go straight to it by using the bookmark dialog box.

Cross-references can also be added to bookmarks. For example, after inserting a bookmark in a document, a cross-reference can be created that refers to that bookmark from other places in the text.

Bookmarks are powerful and a much-underutilized feature in Word, but like Styles, bookmarks can be named however the user or firm desires, and may disclose the origination and originator.

Another issue with bookmarks is that they may disclose the unintentional production process. For example, many document comparison products that do not leverage Word's native tracked changes use

their own process for marking revision changes. This process is in the form of bookmarks. For deleted text, the bookmark name might be XYZ_DeletedText and for inserted text, the bookmark might be XYZ_InsertedText. Unintentionally leaving bookmarks like this in a document may disclose the revision process.

As with custom styles, keep it simple. Use a generic naming convention when naming bookmarks. For example, instead of naming a bookmark "SMHAttorneySignature," it could be more generic with "bkAttySig." The prefix "bk" is an industry standard for bookmark prefixes. Do not use a firm or a person's initials in the bookmark name. More importantly, do not create hidden bookmarks through automation. Hidden bookmarks begin with an underscore ("_"). They are very difficult for users to see and manage, therefore making it easy to unintentionally disclose subtle metadata.

When documents are leaving the electronic walls of a firm, set the firm's metadata management application to remove all bookmarks from documents.

Document Variables. Document variables are a collection of variable objects that represent the variables added to a document or template. They are used to preserve value settings for future reference and are added through programming automation. Without programming knowledge, users cannot add document variables.

While users cannot add document variables, applications that are integrated with Word can and do. Document variables are insidious. They stay in the document like a bad fungal infection. Users have no idea they exist or what they disclose. Even Word's Document Inspector ignores them. The following are document variable examples I've seen that disclose important metadata:

- Client matter information (number and name) that has no relevance to the current document;
- Server path information from the original WordPerfect document that was converted;
- Dates and times of specific document production events; and
- Author information.

I advise not using document variables. If your firm has macros or automation that adds document variables to all documents, use another way of storing that information, such as custom document properties. Also, provide a state of the art document metadata viewer to view all document metadata, including document variables.

When documents are leaving the electronic walls of a firm, set the firm's metadata management application to remove all document variables from documents.

Template Name. Every document originates from a template, and it is easy to determine from what template a document was created. Firms are notorious for giving their templates unique firm specific names. Just like style names and bookmarks, the names may disclose origination and originator.

Use generic naming conventions for all firm templates. Do not place the firm's name or initials in the template name. When documents are leaving the electronic walls of a firm, set the firm's metadata management application to set the document template to be "normal."

Custom Properties. Similar to document variables, custom properties are a collection of document properties that represent the variables added to a document or template. Custom properties are used to preserve value settings for future reference. The biggest advantage of custom properties over document variables is that they are visible and the user can delete or modify them. Nevertheless, like document variables, they stay with the document and can disclose metadata the user may not want others to see.

Attorneys should use generic naming conventions for all custom properties and not place the firm's name or initials in the custom property name.

When documents are leaving the electronic walls of a firm, set the firm's metadata management application to remove custom properties.

SUMMARY

For discovery purposes, it is important to know what metadata is in your documents and what it discloses. In addition to the hard metadata (common metadata elements), there is subtle metadata. Subtle metadata is inconspicuous, but to the trained eye may disclose information such as origination, originator, and the editing process.

To ensure that the subtle metadata is managed properly in your documents I recommend a using one of the document metadata management solution applications available on the market.

Randall Farrar is the president and co-founder of Temecula-based Esquire Innovations, Inc. (www.esqinc.com), a software company that develops Microsoft Office integrated applications for the legal market. He can be reached at Randall.farrar@esqinc.com

"Reprinted with permission from the March 7, 2011 issue of The Recorder. © 2011 ALM Media Properties, LLC. Further duplication without permission is prohibited. All rights reserved."