

Microsystems Engineering Company
2500 Highland Avenue Suite 110
Lombard, IL 60148
Phone 630-261-0111
Fax 630-261-9520
www.microsystems.com



“...Shares Well with Others...”

Coping with “Metadata” Issues

June 6, 2000

Contents

Welcome to the Wonderful World of Document Sharing	1
What is Metadata?.....	1
More Than Metadata!	1
A Big Surprise?.....	1
<i>What's In The Document's Template?.....</i>	<i>2</i>
Who Created The Template?	2
What's In The Header/Footer?.....	2
How Can Our Styles Identify Us?	2
<i>Are These Items In My Shared Document?.....</i>	<i>3</i>
Comments.....	3
Revision Marks	3
Versions	3
Hidden Text	3
Hyperlinks.....	4
Linked Objects/Images	4
Appendix.....	5

Welcome to the Wonderful World of Document Sharing

How did we live just a few short years ago, before it was easy to share our documents with our colleagues, our clients, complete strangers? The truth is, this progress, like so much else, is a double-edged sword. What we're learning is, when you share documents, you share everything about those documents, the good, the bad, the ugly!

Lately the big buzz word is "metadata." We all know we have it. We've all been told we shouldn't share it. But how many of us can even define it?

What is Metadata?

A dictionary definition of metadata is not terribly satisfactory: The word "meta" derives from the Greek, meaning (among other definitions), "with, among, after." In the information technology world, *meta* is a prefix that means "an underlying definition or description." In words related to data and information, the prefix carries the meaning of "more comprehensive or fundamental." Thus, *metadata* is a fundamental definition or description of data.

Let's dig a bit deeper. Certainly, metadata is electronic document information, sometimes embedded where a user can't get at it. In the Appendix at the end of this document you'll find a table listing resources for information about electronic metadata and how to clear it.

More Than Metadata!

But isn't the problem of sharing files really more pervasive even than that? (After all, if I don't give you my electronic document, I don't have to worry about Metadata!) At Microsystems, we work daily with our client's files. Often we can glean huge amounts of information about the document—who probably created it, where it came from, how many places its been, and so on. Even from a printout you can sometimes tell who wrote a document, based on nothing more than an author's "style."

Our focus is this other, less-publicized, but equally important tell-tale data you may not wish to share with others. Think of file sharing as the "document swim-suit competition." Let's be sure our documents only display what we want them to!

A Big Surprise?

It should come as no surprise that sharing documents with clients, co-counsel, and the courts raises an ugly set of potential problems. If you swap electronic spit with someone, you ought to be aware of exactly what you might end up sharing! Here are some items you may want to think about before you attach that file to your next email.

What's In The Document's Template?

A template is a collection of instructions Word is supposed to follow in order to create a set of documents which resemble each other, and follow the same rules. Templates are extremely individual, and vary widely between firms. Their contents reveal much detail about where they were created and what they are intended to do:

Who Created The Template?

There are a number of companies which are hired by firms to create and customize templates. Alternatively, firms may make use of internal resources to create their templates. No matter who creates them, a common technique is to create a collection of document variables, which can be as individual and identifiable as fingerprints! Their presence can serve to identify a document's origin and it's path through the legal community.

Both DocXamine and DocXtools can search for and report on those document variables where they exist. They can be removed through automation, or by moving the document content to a new file container.

What's In The Header/Footer?

Internally, firms want to make it easy to retrieve and edit a file, so they automatically place the filename, the version, the path, the date or time it was last modified, or other information deemed important, into the document's footer. This data is frequently overlooked when the document is shared. It can be removed through automation, or you can delete it manually.

How Can Our Styles Identify Us?

In an effort to make understanding and using styles easier, or to force it's styles to remain together in a style list, or via the use of third-party software, firms often create custom style names that contain the firm's initials, or some other identifying text.

Further, it's common for shared documents to acquire additional styles from every location in which it is edited. For knowledgeable eyes, this is a detailed road map of the document's journey through life.

The use of special or unusual formats in styles can sometimes identify the origin of a document. If a client uses Bookman Old Style and you don't, another client receiving a document from you in BOS could be pardoned for being a tad suspicious of its origins.

Take time to determine how your style names will impact the document when it is shared. Check the document's style list, and use *Format/Style/Organizer* to delete unneeded/unwanted styles.

Are These Items In My Shared Document?

Often a document's final form has little in common with its first draft. We collaborate in the creation of a legal document, and use several Word features to assist us in the process of crafting it. Although not truly "metadata," many of them can have serious repercussions if they make their way outside the firm.

Comments

Comments can contain text which may be inappropriate to share. People sometimes use language during document creation that they would never want in the final version. Also, unless it was removed from *Tools/Options/User Information*, typically your name is associated with any comment you make.

Right-click in a comment and choose *Delete Comment* to remove it. To remove all the comments in a document, use *Browse by Comment* on the Browse Object to move to each, then delete it. (Or, of course, you can write a macro to perform this task!)

Revision Marks

When collaborating during document creation, firms often use *Track Changes* to mark changes to a document, and track who has made them. Ultimately, changes must be either accepted or rejected in order for a document to be truly "final."

Accepting or rejecting changes (in *Tools/Track Changes/Accept or Reject Changes* removes all traces of the track changes feature.

Versions

Even firms which use a third-party document management system like PC DOCS Open or iManage can be tripped up by the versioning feature available in native Word. Once your document is no longer under the control of your DMS, anyone can make use of Word's *Versions* feature, which allows you to save multiple versions of the same document within the same file.

DMS users may assume that no other document versions exist. Before sharing any file, check for Word versions under *File/Versions*. Select the version(s) to be removed, and choose *Delete*. Close the Versions dialog box, then choose *File/Save* to force Word to accept the deletion.

Hidden Text

Occasionally people include information in a document in the form of hidden text. Hidden text is often somewhat offhand, as for example, "Sue—don't forget the blind copy for Joe Smith!" Sharing these casual notes can be embarrassing and unprofessional.

Paradoxically, to remove hidden text, you must first display it. In *Tools/Options/View*, be sure either “Hidden” or “All” is checked under “Formatting Marks.” Next, use *Edit/Replace* to search for and replace all instances of the font format of “Hidden.” Finally, turn off the display of hidden text by unchecking “Hidden” or “All” in *Tools/Options/View*.

Hyperlinks

Documents can contain hyperlinks to other documents or Web pages on an intranet or the Internet. They are identifiable as blue underlined text. These hyperlinks could maintain a connection to a site you may not wish to disseminate, or the hyperlink may not work if it is to a document which is behind a firewall. The bottom line is that it’s a bit “untidy” to share a file under these circumstances.

To remove a Hyperlink, right-click it, choose *Hyperlink* on the shortcut menu, and then choose *Remove Hyperlink*.

Linked Objects/Images

If a document contains linked images or other objects, the path to the linked image or object could be contained in the field code. You can remove linking information by editing the field code.

First, make the field codes viewable. Choose *Tools/Options/View* and check the *Field codes* check box. Next, browse to the field codes in the document and see whether any of them contain linking information. (Press Alt+F9 to toggle between the field code and the image or object.) Select the linked image/object or it’s field code and press Ctrl+Shift+F9. (Note that unlinking an image or object may make it uneditable in the future.)

Appendix

Other Resources

Metadata information specifically related to Word 97 and Word 2000 documents:

Microsoft TechNet	Q223790 Word 97 Q237361 Word 2000	How to Minimize Metadata in Word Documents
	Q71999 Word 97	How to Disable the Fast Save Option in Word for Windows
	Q190733 Word 97 Q211209 Word 2000	Opening Word Document in Text Editor Displays Deleted Text
	Q192480 Word 97 Q197978 Word 2000	Frequently Asked Questions About “Allow Fast Saves”
	Q195005 Word 97 Q195007 WD2000	Some Document Properties Populated Automatically
	Q216866 Word 97 Q194606 WD2000	Summary Information Under Properties Is Not Encrypted
	Q178121 Word 97 Q209638 WD2000	No Password Prompt for “Modify” in Mail Client
	Q194494 Word 97	Password Protection Lost When Saving as Previous Versions
	Q170940 Word 97	Password Not Prompted with First Expansion of Sub-Document

Metadata information or specific solutions:

Perfect Access Software Training & Consulting	http://www.perfectaccess.com/22.htm	Minimizing Word Metadata Risks
The Legal MacPac	http://www.legalmacpac.com	
Payne Consulting Group, Inc.	http://www.payneconsulting.com	MetaData Assistant
TechLaw, Inc.	http://techlawinc.com	
Softwise Consulting Services, Inc.	Phone: 718-871-9776 email: sales@SoftWise.net	

Metadata information extended beyond the Word document:

<u>DocBook, The Definitive Guide</u>	Norman Walsh & Leonard Mueller, Authors	O'Reilly & Assoc., Publishers
<u>The XML Handbook</u>	Charles F. Goldbarb & Paul Prescod, Authors	Prentice Hall PTR, Publishers
<u>XML for Dummies</u>	Marvia H. Aviram, Author	IDG Books Worldwide, Publishers
<u>XML by Example</u>	Sean McGrath, Author	Prentice Hall PTR, Publishers

Acknowledgments

With thanks as usual to Sherry Kappel,
whose original idea fueled this pursuit

Copyright © 2000 Microsystems Engineering Company

All Rights Reserved