

This Free E-Book is brought to you by Natural-Aging.com.

100% Effective Natural Hormone Treatment
Menopause, Andropause And Other Hormone Imbalances
Impair Healthy Healing In People Over The Age Of 30!

What Is Metadata And Why Is It So Dangerous?

By Michael Goldstein, Esq.

Metadata by definition is data about data. It leaves an electronic trail of all the changes to an electronic document: the edits, authors, times and dates of access and changes, as well as routing information. Meta data can tell you:

· Who said what? · Who knew what? · Who did what? · How much time was spent editing the document? · Who created the document? · Where the document has been? · How many times the document was edited? · If and by whom the document has been printed? · Private comments and prior drafts of the document

Metadata is hidden information embedded in many electronic applications, including WordPerfect, Microsoft Word, PowerPoint, and Excel. This data can contain sensitive information, which you would not want disclosed to third parties. Much of this information is invisible to routine users, is created automatically and cannot be prevented.

Even if a document is password-protected, a significant amount of metadata can be discerned even though the document itself cannot be opened. A civil litigation lawyer can use metadata as an effective cross examination tool.

Why civil litigation lawyers and accountants want to know about metadata:

Civil litigators are increasingly sending documents by e-mail during negotiations and discovery. More civil litigation lawyers are realizing the vulnerability of sharing electronic data with clients, opposing counsel and courts.

Your firm can aggressively use metadata on opposing counsel and their clients and others can use it against you. This puts your client relations in danger. By inadvertently revealing private information contained in metadata, you would breach attorney client privilege.

The most common way to protect against metadata is to not copy and paste from one document to the next. Make sure "turn off fast saves" is selected in Microsoft products. Make sure to either accept or

What Is Metadata And Why Is It So Dangerous?

reject all changes to a document, if "track changes are turned on." It is also a good idea once your document is complete, that it be converted into a PDF document. PDF files retain much less metadata than other document forms, especially those created by Microsoft. You can convert a document to a PDF by using either the Adobe Acrobat software, or any number of free online and stand-alone applications that can be found on the web, such as

<http://www.expresspdf.com/>

.

The article was written by Attorney Michael Goldstein, a

<http://www.goldsteinandclegglaw.com>

, of the

Law Office of Goldstein & Clegg, LLC.

SEO Your PDF's

By Kevin Kantola

SEO Your PDF's by Kevin Kantola

Why optimize?

First, why would anyone want to search engine optimize their PDF files? Well, if you had an eBook, brochure, product description or technical document in PDF format, you may wish to optimize these to pick up some extra search engine traffic.

Can the search engines read PDF files?

Yes, most of the major search engines now can read the basic contents of PDF files, though getting these pages to rank as well as HTML files is still questionable.

How is it supposed to work?

This is how the workflow is supposed to work. Create your file in MS Word, or in a draw or page layout program that later can be distilled into a PDF (with some applications you will have to create an EPS file first and then distill it and with other applications, you can distill right out of the apps). If you are using a program such as MS Word, be mindful to apply the H1, H2,

What Is Metadata And Why Is It So Dangerous?

H3 tags where necessary and optimize the body text as you would an HTML file.

When you are finished, distill the file. Bring this file into the full version of Adobe Acrobat 6 for editing. Plug in the appropriate content, post the PDF on your website and let the search engine robots index the file.

How do I plug in the appropriate content?

In Adobe Acrobat 6 there are two places to input content into a PDF file. The first place is under File / Document Properties and the second place is under Advanced / Document Metadata. Under File / Document Properties there are several menus but the most relevant for our purposes is the Description menu. Under the Description menu, there are fields for Title, Author, Subject and Keywords.

Now to confuse matters more, let's go over to the Advanced / Document Metadata menu. There are a couple of choices here, but

let's once again look at the Description menu. Under this Description menu, there are fields for Title, Author, Description, Description Writer, Keywords, Copyright State, Copyright Notice and Copyright Info URL.

How does the PDF store the data?

With duplicate fields, it is important to find out how the data is stored so that we may make some educated guesses as to how the search engines read this data. I performed a few small experiments and here is what I have found. The Title and Author fields seem to be linked to each other because when you change one and check on the other you will see it too has changed. Also, the Subject field of the Document Properties menu seems to be linked to the Description field of the Document Metadata menu for the same reasons. The Keyword fields, however, are not linked. Separate sets of keywords can be added to both fields. When the file is saved, both sets of keywords are stored in the PDF file.

Which set of keywords is correct then?

Adobe stores its metadata in XML format. Opening the PDF file in Notepad, it appears that the Keyword field under Document Properties is the one that the search engines will use (this

What Is Metadata And Why Is It So Dangerous?

hasn't been proven, yet though). The keywords input into this field appear in the PDF as we have come to expect, separated by commas, like this: Keywords(movies, cinemas, matinees, theatres, popcorn).

The keywords that were input into the Document Metadata menu appear as a sort of list like this:
treeswoodchips

Of course, this doesn't mean anything really - it is how the search engines read this that counts.

How does it really work?

I've run some preliminary tests (and by this I mean very preliminary) and more testing will need to be completed to verify these results, but here is what I have come up with so far. When a PDF file was first opened in Acrobat 6 the Document Properties or Document Metadata title and author fields were already filled in with the file name and author's initials (information received from MS Word)

Without filling in any extra data into the Document Properties or Document Metadata menu, Google used the Title field information

for the title in the results and the description in the results was acquired from the body copy. Yahoo!, in older PDF's use the largest text on the page as the title text. In regards to more recently indexed PDF documents, however, Yahoo! is using the Title field information as the title text in the search results. At this writing, the description text in the search engine results comes from the body text of the PDF and not the Document Properties or Document Metadata text.

Thinking I might just get lucky (and hoping for quick results), I ran a few optimized and non-optimized PDF's through some of the more popular search engine spider simulators on the web, but these spiders did not handle the binary code very well. None of them returned title or meta tag information and the most popular keywords were snippets of binary code.

So, at this point, does it really pay to optimize a PDF?

The simple answer is, yes. The title tag and body copy can still be optimized and the major search engines will index it

What Is Metadata And Why Is It So Dangerous?

accordingly. As far as the Keywords and Description meta tags, well Google ignores this in PDF's just as it does in HTML documents and Yahoo!, which does use the description tag, is only half way to where it needs to be.

But Google and Yahoo! aren't the only two search engines / directories around and with algorithms changing all the time, perhaps someday soon either the SE's will be able to fully read a PDF file or Adobe will offer a patch that will make PDF's more SE-friendly. It's only a matter of time, my friend. Will you be ready?

Copyright © 2004 SEO Resource <http://www.seoresource.net> Kevin Kantola head's up SEO Resource, a California search engine optimization company devoted to achieving high rankings.



This Free E-Book has been brought to you by Natural-Aging.com.

**[100% Effective Natural Hormone Treatment](#)
Menopause, Andropause And Other Hormone Imbalances
Impair Healthy Healing In People Over The Age Of 30!**

What Is Metadata And Why Is It So Dangerous?

