

Document Properties Document Statistics & File Dates
Document Reviewers Custom Properties **Track**
Changes and Document Revisions Header
and Footers **Footnotes White Text Small**
Text Macros Previous Versions Routing Slips
Fast Saves Hidden Slides Hyperlinks



Guide to the Dangers of Hidden Information in Documents

A Workshare Report
Published 2011

© 2011 Workshare, Inc. All rights reserved

Guide to the Dangers of Hidden Information in Documents

Executive Summary

Hidden information in documents can pose a serious risk, yet many people are not even aware of the danger. A potent example is document metadata—hidden information contained in Microsoft® Office documents, including Microsoft Word, Microsoft Excel, and Microsoft PowerPoint files. Whenever a document is created, edited, or saved, metadata is automatically added to it at the system and application level. Microsoft Word's popular collaboration features, such as comments and track changes, result in a significant amount of metadata being included in documents through actions of the author. This information can be transmitted every time a document is emailed to someone either inside or outside the organization.

Metadata was originally conceived to make it easier to track and find document data. When used properly, metadata is useful. But when used carelessly or ignored, it gives unauthorized people access to privileged information that could be used against you or your organization.

What is the Risk?

The problem is not that metadata is added to a document, but that it is difficult to fully identify and remove. For example, in Microsoft Word, the ability to add comments and track changes can be very helpful to users collaborating on a document. However, 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. This can result in embarrassing situations where external parties see information that was not intended for their eyes.

Metadata can put you at financial risk, a competitive disadvantage, or in an embarrassing situation with costly consequences.

Metadata risks are further compounded when documents are attached from within Microsoft Outlook and sent to outside parties. Used as the de facto way to electronically exchange documents in an enterprise environment, Microsoft Outlook does not offer warnings about metadata in attached documents or zipped files. Thus, the potential to accidentally send documents containing harmful metadata (and thus expose an organization to the inadvertent disclosure of sensitive information) is amplified tremendously with every document that is sent back and forth during the collaborative process.

Metadata accidentally left in documents can be easily viewed by anyone who has access to these documents. Because metadata is not always viewable, document users can unwittingly send confidential information to people outside their organization. In fact, there have been several widely publicized, high-profile cases in which document metadata proved to be the culprit. The bottom line is: document metadata can get organizations in big trouble—by putting the organization at financial risk, a competitive disadvantage, or in an embarrassing situation with costly consequences.

Guide to the Dangers of Hidden Information in Documents

High Profile Cases of Metadata

Metadata can often pose risks to an organization when the information that is revealed is meant for internal use only. This information can be anything from revealing confidential statistics, to exposing proprietary comments from document reviewers. Sharing this information, or simply leaving it in the documents, can have unintended consequences. Below is a list of high profile cases involving the metadata in documents.

- **Google:** Google revealed private financial forecasting when hidden data was left in a PowerPoint presentation before posting it for the Wall Street community
- **Microsoft:** Through hidden data within Microsoft Office documents, the Associated Press found that Microsoft's advertising campaign highlighting a customer that switched from Apple to Microsoft's software was in fact a member of their PR firm
- **Whole Foods:** Court documents containing hidden information disclosed Whole Foods plans to close stores, revealed how Whole Foods negotiates with suppliers to drive up costs for Wal-Mart, and disclosed the company's closely held marketing strategies
- **Barclays:** An Excel spreadsheet contained 179 contracts within hidden columns that were then accidentally submitted in Barclays buyout offer of Lehman Brothers assets
- **Google:** Hidden metadata revealed Google as the submitter of a PDF document to the Australian Competition Commission and Consumer Commission (ACCC) protesting the removal of all payment options except PayPal by eBay Australia
- **AT&T:** AT&T revealed confidential information about spying on their customers when a PDF file was released that included hidden information
- **Telxon Corp.:** Abrupt deletion of all metadata in two years of documents was not considered to be in "good faith" production and resulted in a default judgment as sanction
- **Alcatel:** A security vulnerability in Alcatel's DSL modems was revealed in document metadata
- **SCO Group:** This leading software provider filed a suit against DaimlerChrysler and AutoZone and track changes left in a Microsoft Word document revealed that considerable time was spent focusing on Bank of America as the defendant and not the automaker
- **Westpac:** The oldest bank in Australia revealed a full-year of profit results via metadata before it was finalized and lodged with the Australian Stock Exchange

Guide to the Dangers of Hidden Information in Documents

- **Merck:** Metadata revealed this global healthcare leader deleted vital information concerning the arthritis drug, Vioxx, which resulted in users having false information on heart attack risk associated with taking the drug
- **Sun Life Financial:** Hidden data found in a document, forced the company to release its fourth-quarter and year-end results ahead of schedule
- **General Electric:** Information on a sex discrimination case involving GE's management was revealed through PDF files exchanged by lawyers working on the case
- **SCO Group:** In 2003 SCO filed a complaint that included hidden track changes and comments that revealed its legal strategy and raised issues of concern
- **Eli Lilly's and Co:** Eli Lilly's lawyers accidentally emailed documents that contained confidential information to a reporter for the New York Times
- **The British Prime Minister's Office:** Hidden data in the UK's government "Dodgy Dossier", the document that helped propel the country into war, revealed a student paper was the source of the document
- **Department of Justice:** Through the discovery of data hidden in documents, it was revealed that Barry Bonds, among other professional athletes, were involved in the BALCO performance enhancing drug scandal
- **The United Nations:** Metadata revealed the UN office doctored a report on the murder of the former Lebanese Prime Minister, Rafik Hariri
- **Democratic National Committee:** Judge Sam Alito inadvertently revealed his true beliefs on immigration laws and other issues when memos were released containing blacked out data
- **Scottish Council:** Incriminating information on the Scottish council was revealed when tracked changes were accidentally left in a document version of a waste management report
- **National Intelligence Budget:** Detailed hidden data in a PowerPoint document posted online by the Office of the Director of National Intelligence revealed the confidential National Intelligence budget
- **Australian Government Ministers:** Left over hidden information in a report revealed author names and caused Australian government ministers to resign

Guide to the Dangers of Hidden Information in Documents

- **UK's Civil Aviation Authority:** "Restricted and confidential information" was revealed when a document by the UK's Civil Aviation Authority was posted on their website that contained hidden data
- **Justice Department:** A PDF file released by the Justice Department revealed Social Security numbers
- **UK Anti-Terror Bill:** In a Word document, the former Home Secretary accidentally leaked the government's indecision about new anti-terror laws to members of the opposition, leading to a speedy defeat of the bill
- **Pentagon:** In a PDF file with hidden information, the Pentagon revealed confidential details about the death of a US agent
- **Washington Sniper:** Confidential telephone numbers were revealed by uncovering hidden information in a report on the Washington Sniper
- **FCC Chairman:** By leaving hidden information in his letter supporting tougher regulation on broadband, Rep. Jay Inslee revealed a lobbyist's letter was the original author of his letter

Types of Document Metadata and Their Associated Risks

Document metadata comes in many forms. Below is a list of the types of metadata found in Microsoft Office documents and the risks that each type poses:

Document Properties

Applies to: Microsoft Word, Microsoft Excel, and Microsoft PowerPoint documents.

Document properties are details about a file that help identify that it includes a descriptive title, subject, author, manager, company, category, keywords, comments, and hyperlink base.

Document properties display information about a file to help organize the files so that they can be easily found at a later date.

Risks: The names of authors and the name of the organization can display sensitive information. It is possible that if a document has been sent outside your own organization, the author name and company name contained in the built-in properties could be a name other than your own. In addition, if documents are re-purposed or used as a template for a new document, information that is specific to a previous client such as pricing, terms, or the client's name can be stored as hidden information within the new document.

Guide to the Dangers of Hidden Information in Documents

Document Statistics & File Dates

Applies to: Microsoft Word documents only.

Document statistics include information on when the document was created, modified, accessed, and printed. In addition, document statistics display the name of the person it was last saved by, the revision number, and the total editing time. Other statistics include number of pages, paragraphs, lines, words, and characters.

Risks: Document statistics can create embarrassing situations. For example, a law firm might bill for more hours than the document's total editing time. In addition, the "last saved by" metadata shows the last person who edited the document and can create discrepancies over who worked on a document.

Document statistics can reveal more hours than the document's total editing time has been billed

Document Reviewers

Applies to: Microsoft Word documents only.

Document reviewers consist of a list of users that have added or accepted document changes.

Risks: Document reviewers metadata can expose who has suggested what changes. Removing the names of reviewers can be as important as removing the changes they have suggested.

Custom Properties

Applies to: Microsoft Word, Microsoft Excel, and Microsoft PowerPoint documents.

Custom Properties includes any property fields added manually to a document or by various programs to help manage and track files.

Risks: Custom Properties are normally specific to an organization. Common types of custom properties are document ID, department and status. Custom properties can reveal proprietary information or competitive business practices.

Hidden Text

Applies to: Microsoft Word documents only.

Hidden text is text blocks that have been formatted as hidden. Unless specifically selected to be viewed in Microsoft Word, hidden text is not displayed within the document.

Risks: Hidden text can contain sensitive information that unauthorized parties may potentially view.

Comments

Applies to: Microsoft Word, Microsoft Excel, and Microsoft PowerPoint documents.

Comments are notes and suggestions that are added to a document via the comment feature to help facilitate an online review.

Guide to the Dangers of Hidden Information in Documents

Risks: Comments, like hidden text, can display sensitive information to external parties because comment metadata travels with the document. Microsoft Excel and Microsoft PowerPoint documents are especially susceptible to this risk as there is no internal mechanism built into these applications to warn a user that comments are embedded. Gartner Group states that “while Microsoft is aware of potential problems [with comments], it does not have a comprehensive solution to solve this problem.”¹

Track Changes and Document Revisions

Applies to: Microsoft Word and Microsoft Excel documents.

The track changes feature tracks text that has been inserted, deleted, or moved during an online review. As changes are made to a document, a new revision of the document is kept by the application. This revision history exists even after changes to the document have been accepted or rejected.

Risks: If track changes is left on, every change made to the document is recorded. This is like recording every single keystroke made to the document that can be viewed by subsequent reviewers. Thus, even though the changes may not be visible using “Final” review mode, they still travel with the document and can therefore be seen by unauthorized parties with potentially disastrous consequences for your organization.

Headers and Footers

Applies to: Microsoft Word, Microsoft Excel, and Microsoft PowerPoint documents.

Headers and footers are areas in the top and bottom margins of each page in a document. Text or graphics such as page numbers, the date, a company logo, the document’s title, file name, or the author’s name can be inserted there.

Risks: Custom headers and footers can contain descriptions such as filename, path, the date and time the document was modified, or other information that makes it easy to retrieve and edit a file. Unfortunately, the information contained in footers and headers is often overlooked when the document is shared, potentially exposing confidential information.

Footnotes

Applies to: Microsoft Word documents only.

Footnotes attributed to content are embedded as metadata into Microsoft Word documents.

Risks: Footnotes may expose private, internal documents used within the organization or give access to private data via hyperlinks.

White Text

Applies to: Microsoft Word documents only.

White text is blocks of text that have been formatted with a font color of white on a background of white. The text appears invisible when viewed or printed and can be used to hide information in a document.

¹ Silver, Michael, Metadata in Office, Gartner Group Research Note, 23 January 2003

Guide to the Dangers of Hidden Information in Documents

Risks: White text is commonly used when documents are posted to the Internet so they can be more readily found by search engines. However, white text can also be viewed by external users and, depending on what was actually written as white text, it can be damaging. White text can also be used for field codes such as the “include text,” which points to a file location. If this file location code is embedded in a document, users can unknowingly update it and share it with people outside the organization.

Small Text

Applies to: Microsoft Word documents only.

Any text block in a document that is less than five points is considered small text. The text is so small that it is not visible when viewed or printed, and can be used to hide information in a document.

Risks: Like white text, small text is commonly used to put information in documents so they can be found by search engines. Small text can also include sensitive information that was not meant to be distributed externally.

Macros

Applies to: Microsoft Word documents only.

If a task is repeated in Microsoft Word, it can be automated using a macro. A macro is a series of commands and instructions that are grouped together as a single command to accomplish a task automatically.

Risks: There are several reasons to strip out custom macros. For example, macros can be set for templates that may have some amount of prepopulated data. There may be a time when the information contained in these templates should not be seen by external audiences. Macros can also be linked to internal databases or intranets. Having access to the internal file naming structure is generally information that most organizations do not want outside their firewalls. Lastly, macros are often quite complex and, if developed in-house, may represent the company’s intellectual property. If macros are included in the document, the information is freely shared with any outside party.

Previous Versions

Applies to: Microsoft Word documents only.

Previous versions show the number of times that a document has been versioned over its lifetime. This function enables Microsoft Word to save prior versions of a document as a part of the electronic file.

Risks: The risk associated with previous versions is that a recipient can access any of the previous versions that have been saved. Therefore, the party reviewing the document can go back to any version and see what was changed in the document lifecycle. This metadata, while useful in some instances, can disclose sensitive information.

Routing Slips

Applies to: Microsoft Word and Microsoft Excel documents only.

Guide to the Dangers of Hidden Information in Documents

Routing slips are used to create a distribution list of reviewers in a particular order. Routing slips are manually created by adding in recipients' email addresses. When files are routed, routing slips are sent as email attachments.

Risks: Routing slips reveal the names of document reviewers. This may be information that should stay confidential rather than distributed externally. An example of how this information can be used is when email addresses are put in the routing slips. If this document is then published to the Internet, the email address can be displayed for all to see.

Fast Saves

Applies to: Microsoft Word documents only.

Fast saves is an option that saves just the changes that were made to a document, resulting in the history of document changes. Turning off fast saves and saving the document the standard way removes the changes and stores only the final version of the document.

Risks: Like other metadata, changes saved during a fast save can expose sensitive information to external parties when viewed using a text or hex-editor. Deleted text can still exist in the electronic file. According to the Gartner Group's Research Note on Metadata in Microsoft Office, "users can easily forget that metadata exists when they send the document to someone else. Some metadata is never visible, such as pieces deleted by users but not really deleted by Microsoft Office when operating with fast save turned on."²

Hidden Slides

Applies to: Microsoft PowerPoint documents only.

Hidden slides are slides that are not displayed during a slide show.

Risks: A master Microsoft PowerPoint slide deck may contain some slides that are used as backup or that are for internal use only. To prevent accidental showing of these slides, it is best to strip out any hidden slides before sending the slide deck out externally.

Hyperlinks

Applies to: Microsoft Word and Microsoft Excel documents only.

Documents can contain hyperlinks to other documents or web pages and are displayed as blue underlined text. Hyperlinks in Microsoft Excel files can be seen in: a link to a cell in another Microsoft Excel document; a named link to a named reference in another Microsoft Excel document; a link to another document; an OLE link that inserts another document as an icon; and an OLE link that inserts another document as text.

Risks: Hyperlinks can maintain a link to a site that organizations may not wish to disseminate, such as files that may exist on a computer's local file system, on an organization's internal database, or on an intranet. Disclosing the file path, or the location of where the files are stored, can invite potential hackers to gather sensitive information.

² Silver, Michael, Metadata in Office, Gartner Group Research Note, 23 January 2003

Metadata Legal Opinions

The handling of metadata has created conflicting legal opinions. Some suggest that metadata should be available to the public, while others stand by the idea that it should be the owner's choice as to whether they should make metadata public information or not. Below are some examples of these conflicting views.

- **American Bar Association Ethics Opinion:** Lawyers receiving electronic documents are free to examine hidden metadata.
- **Nova Measuring Instruments Ltd. v. Nanometrics, Inc.:** Held that metadata should be produced
- **Wyeth v. Impax Laboratories, Inc.:** Held that production of metadata is not required absent a strong showing of particularized need
- **New Amendments to the Federal Rules of Civil Procedure:** Affect data retention policies and electronically stored information
- **AstraZeneca:** In a products liability suit involving the drug Seroquel, a federal court ordered specific metadata fields must be produced
- **The Alabama State Bar Commission:** Believes that it is acceptable to hide metadata if this means that clients' secrets are kept confidential
- **The Arizona State Bar:** Stands by the notion that it is the duty of the lawyer to protect his or her client by making sure that metadata does not fall in the hands of the wrong person
- **The Vermont Bar Association:** Has found nothing that compels them to believe that lawyers who receive electronic files do not have the right to use certain tools which expose metadata
- **The West Virginia Bar Association:** Does not believe that lawyers have the right to view metadata that is meant to be hidden
- **New York State Bar Association:** Opinion brief recommends that attorneys remove metadata from documents sent via email

Guide to the Dangers of Hidden Information in Documents

- **Big Pond Communications v. Kennedy:** 70 O.R. (3d) 115 held that the path-and-filename metadata found at the bottom of the statement of claim, even though irrelevant, came within the ambit of the pleading, was privileged and was therefore not actionable.
- **The Sedona Guidelines:** “Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age” suggests that “Absent a legal requirement to the contrary, organizations are not required to preserve metadata; but may find it useful to do so in some instances”

Conclusion

Metadata can serve useful purposes for identifying, indexing, and managing documents. It is critical to understand how metadata is created, where it is stored in a document, and how it changes, especially when collaborating. Along with comments, routing slips, macros and other information, metadata in Microsoft Office includes visible text and hidden properties in Word; hyperlinks, and hidden columns in Excel; and hidden slides in PowerPoint. All these types of metadata can reveal confidential information that, if not properly identified and managed, could result in embarrassing incidents, competitive disadvantage, or outright legal action against your organization.

Metadata can and generally should be cleaned before distributing a document. Because metadata is not always viewable, it is possible to unwittingly send a document with confidential information outside the organization. The growth of smartphones and tablets provides the ability to view, edit and redistribute documents via email, further increasing the risk of revealing confidential information outside your system’s firewall.

Although there are conflicting legal opinions about the handling of document metadata, it is clear that this is an issue that deserves attention by every organization producing electronic documentation.

About Workshare

Workshare provides document collaboration software to enable business professionals to accurately create, collaborate and control high-value content more efficiently. More than one million professionals rely on Workshare solutions to increase productivity and safeguard their confidential information, ultimately securing their intellectual property, customer relationships and the organization’s reputation. Workshare’s comparison and collaboration solutions provide risk management and security to over 18,000 organizations worldwide and easily integrate with line-of-business applications. For more information please visit www.workshare.com.