Workshare

A Guide to Managing Metadata in Today's Law Firms

An Osterman Research Report

Author: Osterman Research, Inc | **Published:** July 2010

Executive Summary

There are millions of lawsuits filed in every year. Each of these lawsuits can generate thousands of documents, many of which contain "metadata"—hidden data that consists of notes or other information within word processing documents, spreadsheets, presentations, PDF files and other types of content. Metadata is often used for internal purposes as reviewers collaborate on a document. Rarely is metadata intended for an external audience, particularly when that audience is opposing counsel or the public.

However, there have been many instances where metadata has mistakenly been sent along with the original source documents or posted to a web site. Inadvertent disclosure exposes law firms to the risk of regulatory breach, embarrassment, lost clients and lawsuits—all of which affect a firm's bottom line.

Law firms have an ethical, and in many cases a regulatory, obligation to protect attorney-client privilege by removing metadata from documents presented during discovery or sent outside the firm. This obligation means that law firms must take steps to secure documents that are sent through email, ensure documents are protected from tampering, manage ongoing document security policies, and so forth.

There are a variety of tools available that can scrub metadata from documents, manage corporate policies, monitor violations and provide other capabilities. There are fewer tools that can help educate users to the dangers of metadata, integrate all of the functions into a cohesive system, and provide the legal IT team a central interface to manage metadata across the firm. While point solutions can provide useful functionality for law firms, an integrated solution is less expensive to manage and provides better protection to an ad hoc collection of different vendors' offerings.

ABOUT THIS GUIDE

This guide discusses the challenges associated with cleaning metadata from documents and protecting law firms from the inadvertent disclosure of proprietary information. This guide also discusses the key questions decision makers should ask as they develop risk management policies and deploy technologies to protect confidentiality as it pertains to metadata.

Basic Metadata Protection for Law Firms

JUST WHAT IS METADATA?

In the context of this guide, metadata is information contained within an electronic document that provides additional information about the document itself or certain parts of it, and that generally is hidden from view in the normal display of the document. It is often described as "data about your data."

Metadata can be automatically generated by the application used to create the document, or it can be manually entered by a reviewer, author or commentator. Examples of metadata include:

- A word processing file or spreadsheet that contains information about the author of the document, when it was created, when it was last edited, notes about the document, the number of characters and words it contains, etc. In a Microsoft Word document, for example, some of the metadata about that document can be accessed and modified under File > Properties.

- Documents that include comments and notes from reviewers, auditors and others. Generally, the display of this content can be turned on or off depending on the purpose of the metadata inclusion.

- Improperly redacted information is another common example of hidden document information. The most common mistake is covering text, charts, tables, or diagrams with black graphics, or highlighting text in black, in an attempt to redact information. Quite often the cover up can be removed to reveal the text underneath.

In short, metadata provides critical information about other documents or files, but is rarely intended for display with the primary content included in the document.

WHAT LAW FIRMS SHOULD CARE ABOUT

Electronic documents have been a key component of discovery actions for several years, but the new amendments to the Federal Rules of Civil Procedure (FRCP) that took effect on December 1, 2006 significantly increased the risk for any organization involved in discovery actions by elevating the importance of electronically stored information (ESI). While organizations had been using electronic information in discovery actions for some time, the new amendments essentially codified this practice.

Several state jurisdictions, such as the US District Court of Western Pennsylvania, have put the onus of managing metadata in PDF documents clearly on the submitting parties. The U.S. District Court for the Western District of Pennsylvania issued a warning in February 2010 about the metadata in PDF documents being uploaded to its servers:

> It has come to the attention of the Court that Metadata (AKA Hidden Data) within Microsoft Office Word is not being removed from some documents prior to their conversion to PDF format for uploading to the Court CM/ECF server. Please be advised that Hidden Data can be retrieved from PDF documents if the data is not cleaned from the Microsoft Word document prior to conversion. *February, 2010*

In the U.K., professional rules require all lawyers to keep the affairs of clients and former clients confidential (excluding the provisions of legislation such as the Data Protection Act of 1998). According to Section 2.3, Rule 5 of the Information security practice note (September 11, 2008) principals in law firms must make arrangements for the effective management of the firm as a whole, a responsibility that includes ensuring the documents and assets entrusted to the firm are kept safe.

As a result, the metadata contained within electronic documents raises several important issues on which law firms should focus:

- **Metadata must be managed**
  Metadata must be managed in a coordinated fashion throughout the lifecycle of all electronic documents. Law firms must establish policies about how metadata is created, managed, archived and deleted for all electronic document types. While these policies are influenced by regulatory and statutory obligations for the preservation of data, law firms should be the primary example and source of advice for their clients on creating and maintaining metadata-focused policies.

- **Metadata management is now more complicated**
  The new amendments to the FRCP broadened the scope of the management of metadata in the U.S. While there are no rules in the new amendments or elsewhere that specifically define how to manage metadata, the increased emphasis on ESI also increased the importance of metadata, while offering no new guidelines as to how it should be managed. Further, there are conflicting court rulings in the U.S. and U.K. with regard to metadata:

  - *Nova Measuring Instruments Ltd. v. Nanometrics, Inc.*, 417 F. Supp. 2d 1121 (N.D. Cal. 2006) held that metadata should be produced.

  - *Wyeth v. Impax Laboratories, Inc.*, 248 F.R.D. 169 (D. Del. 2006) held that production of metadata is not required absent a strong showing of particularized need.

  - *Seroquel Products Liab. Litig.*, 244 F.R.D. 650 (M.D. Fla. 2007) held that metadata need be produced only when relevant and not extraordinarily difficult for the party that owns it.

  - Big Pond Communications v. Kennedy, (2004), 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.

- **Metadata represents potentially discoverable content**
  Further, in the U.S. FRCP Rule 26(b)(2) introduced the notion of "accessible" and "inaccessible" data, negating the ability of organizations or their counsel to automatically dismiss metadata as being in the latter category.

CHALLENGES WITH DOCUMENTS

It is important to note that metadata is incredibly useful for codifying, categorizing, annotating, tracking and managing documents. Individual users can manage their documents more efficiently by archiving systems to index content so it has more granularity and is easier to find, while those involved in e-discovery use metadata to find necessary content more quickly.

That said, metadata also poses an enormous risk if it is not managed properly. For example:

- Automatically generated metadata about a document's creation date can reveal that an older document has been repurposed. This can create embarrassment for an organization and lead to problems that could have been avoided if the document had been scrubbed of its metadata.

- Documents reviewed by multiple individuals may contain comments and revisions intended for other reviewers or the original author. The comments might include proprietary information such as confidential product information, off-hand remarks, opinions of individuals that do not reflect the official views of senior management, profanity and other content that clearly should not be sent outside the organization.

- Hidden text and white text both serve useful purposes in the context of managing and indexing electronic documents; however they can contain information that should not be divulged without appropriate review. Further, hidden text and white text can contain links to other documents, allowing unauthorized parties to gain access to internal data stores.

- In most desktop productivity applications, comments can be displayed or turned off. If a document that contains non-visible comments is inadvertently sent outside of the organization, such as to an opposing party during discovery, this can have enormously damaging consequences to an organization and be the difference between winning and losing a case.

- Related to the point above is that some collaboration systems permit multiple reviewers to comment on a document, sometimes without the original author's knowledge. The author, unaware that annotations or other metadata have been added to the document, might send it outside the organization, where the metadata could prove damaging.

- Macros are bits of program code within a document that permit the repetition of certain tasks or that provide enhanced functionality, such as the ability to use buttons in a Microsoft Excel worksheet. However, macros can represent valued intellectual property that an organization might not want to share, or they can point to internal data stores or other proprietary information.

- Other types of metadata include custom properties that represent intellectual property, proprietary content in headers and footers, and small text that is designed not for human reviewers but instead for search engines when a document is placed on a web site.

CHALLENGES WITH PDFs

PDF documents present similar challenges from the perspective of managing metadata. For example, a document converted to a PDF file may include metadata in the form of document properties that allow the document to be indexed more fully by archiving systems, as well as searched with more granularity after the fact. Further, some tools, such as Adobe Acrobat and the Mac OS X Preview function, allow annotations and comments to be stored in PDF documents.

While the likelihood of inadvertent inclusion of metadata in a PDF file is not as high as with other types of electronic files, it can happen. For example, people often convert Word documents to PDF to eliminate comments and tracked changes. However, if these changes are displayed in the Word document when the PDF is created, the changes will also appear in the resulting PDF file. Similarly, if the Print Hidden Text option is selected in Word, hidden text will appear when the PDF file is created.

Risks to Your Law Firm

There are a number of problems that can arise if metadata is not managed properly. Examples include:

- BHP, a multinational mining firm, developed a copper mine in Papua New Guinea that polluted a native fishing area, the Ok Tedi River. A native group brought suit against BHP, shortly after which the government of Papua New Guinea passed a law that made it a criminal act to pursue litigation against BHP for polluting the river. However, the attorneys for the native group found a draft of the legislation in a word processing document that included BHP's footer. BHP was subsequently found guilty of contempt, resulting in a significant fall in their share price. Within two weeks of the contempt finding, BHP resolved the legal action[1].

- In 2006, Google inadvertently showed a presentation during its annual Analyst Day that revealed notes about its projected advertising revenue for the year. The company included its mistaken revelation in a Securities and Exchange Commission filing[2].

- In 2003 a memo was prepared by British Prime Minister Tony Blair's office to support the notion that UN weapons inspections were not working in Iraq and that military action was justified. Richard M. Smith, a privacy and security expert in the US, downloaded the Word document and extracted the metadata. The metadata revealed the press office was deeply involved in the Iraq memo's preparation and copied portions of a US graduate student's work.

- In 2000, Merck submitted an article to *The New England Journal of Medicine* about its anti-arthritis drug Vioxx (which has since been withdrawn from the market after Merck was hit with roughly 7,000 lawsuits). The editors of the *Journal* revealed publicly that metadata showing a link to an increased heart-attack risk with Vioxx had been deleted from the article[3].

- In 2005, the final draft of a United Nations report on the murder of the former Prime Minister of Lebanon omitted critical information about Syria's involvement in the murder. Prior to submitting the report, the author failed to delete the edits[4].

[1] http://tinyurl.com/nqwt9b

[2] http://tinyurl.com/dn85ek

[3] http://tinyurl.com/mkmpsd

[4] http://tinyurl.com/kvo23z

- A contributor to the *Journal of Accountancy* wrote a client an angrily toned memo. After considering his words a bit more carefully, he deleted the original draft and used a more appropriate tone, after which he sent the memo. However, the original version of the memo was included along with the latter[5].

- In April 2005, the Department of Defense Multi-National Task Force – Iraq Unit published a report in PDF format about its investigation of a shooting incident. The authors supposedly redacted sensitive content, but an Italian blogger was able to recover the redacted content using tools available in Microsoft Windows[6].

THE RISKS ARE MORE SERIOUS FOR LAW FIRMS

While the risk of not properly scrubbing documents of potentially sensitive or embarrassing content contained in metadata is serious for any organization, the situation is much worse for law firms for two reasons:

- Law firms, because of the nature of their work, deal with significant quantities of sensitive content. While their clients and other firms outside the legal profession also deal with sensitive content, the quantity and concentration of sensitive content tends to be much higher in law firms. As a result, the potential for mistakenly including metadata in documents that leave the firm is much higher than for the average organization.

- Complicating the issue for law firms is the fact that much of the content with which they work is protected by attorney-client privilege. This places an additional burden on law firms that others do not face, since the consequences for accidental disclosure of metadata in violation of the privilege can be significant.

- For example, Rule 4 of the Solicitors' Code of Conduct (2007), requires that lawyers in the U.K. keep the affairs of clients and former clients confidential, except where disclosure is required or permitted by law or the client. This duty of confidentiality extends to all confidential information about a client's affairs, irrespective of the source of the information.

- Paragraph 702 of the 8th Edition of the Code of Conduct of the Bar of England & Wales goes even further, stating that barristers "must preserve the confidentiality of the lay client's affairs and must not without the prior consent of the lay client or as permitted by law lend or reveal the contents of the papers in any instructions to or communicate to any third person… information which has been entrusted to him in confidence or use such information to the lay client's detriment or to his own or another client's advantage." [emphasis added]

[5] http://tinyurl.com/ksmcyf

[6] http://tinyurl.com/lq89ke

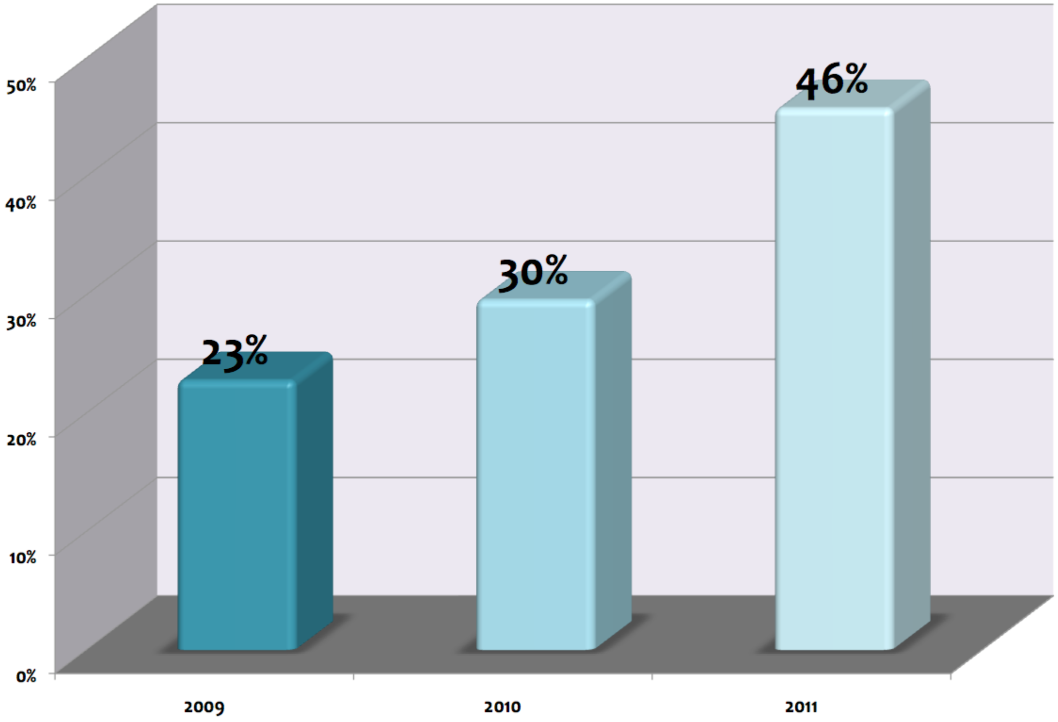
Where is the Industry Heading?

KEY TRENDS IN DATA MANAGEMENT

Among the many trends occurring in data management, there are important trends that impact the management of metadata in significant ways:

- Use of email-enabled mobile devices is increasing at a rapid pace. As shown in the following figure from a survey conducted by Osterman Research in April 2009, the use of mobile devices designed to access email will double between 2009 and 2011.

Proportion of Users that Will Employ Wireless Handhelds to Access Email 2009-2011



What this means in the context of metadata management is that (1) users will have many more opportunities to generate metadata, particularly as mobile phones become more widely used to view and edit documents; and (2) there will be more endpoints from which metadata must be managed. The problem will be complicated by the fact that users can modify documents on their mobile devices and transmit them directly without sending them via a security solution that automatically checks for metadata and/or scrubs it before documents are sent.

- A trend occurring in organizations of all sizes is a rapid increase in cloud-based computing for a variety of computing tasks. A growing number of law firms are migrating to cloud-based desktop productivity applications. This further complicates the management of metadata, since documents are no longer stored on local machines or file servers, but instead stored remotely at third-party data centers.

- Server solutions now include new metadata removal options that protect the entire organization by scrubbing email attachments transmitted via mobile devices and corporate web mail. The downside is that server solutions give end users less control over their documents, but they may still be the preferred solution for environments where end-user control is not a priority. Ultimately, the combined power of today's server and desktop solutions may offer the best overall solution by ensuring end-to-end metadata management along with additional features, enhanced flexibility and better end-user control.

Finding the Right Solution

There are a number of point solutions on the market that address the specific issues discussed in this paper. For example, some solutions focus on allowing the user to scrub metadata from Microsoft Office documents before they are sent in email, while others focus on server based applications that don't rely on the end-user.

It is important to evaluate your organization's needs to determine the right solution for your firm. Beyond the product features, it is also important to consider the following:

- Purchasing from fewer vendors typically makes integration easier and tends to be less expensive per function or feature than many separate, point solutions.

- An integrated solution is easier for IT to manage day-to-day and reduces the amount of training required.

- Similarly, user training is simplified and decreases the complications of firm-wide implementations.

- A solution set from fewer vendors is easier to upgrade than a set of point solutions because the latter will likely be upgraded at different times, and an upgrade in one tool may be incompatible with another.

The bottom line is that the right solution can provide good functionality and address the problems that law firms face in the context of managing their information and preventing inadvertent leaks of metadata.

Summary

Cleaning metadata from word processing documents, spreadsheets, presentations, PDF files and other documents is critical for any organization, but especially so for law firms. Because law firms have a duty to protect their own confidential information and that of their clients, particularly with regard to information covered by attorney-client privilege, they must ensure that metadata is cleaned from outgoing documents, particularly those documents sent to opposing counsel.

Law firms must be able to ensure that violations of legal, regulatory or corporate policy are monitored. Additionally, they must be able to manage all of these capabilities efficiently and with a minimum impact on administrative staff, lawyers, and the legal IT team. Choosing the right solution is very important to ensure that metadata protection is maximized and exposure to data breaches are kept to a minimum.

Product Comparison Worksheet

The following worksheet provides a list of common features and functions to ask for when you review metadata solutions for your firm.

	Solution A	Solution B	Solution C
Metadata Removal			
Are the features available on workstations			
Are the features available as a server solution			
Can you detect and remove metadata in Microsoft Word, Excel, and PowerPoint documents			
Can you remove metadata in PDF documents			
Can you view metadata in Office documents before removal			
Can you clean metadata on Outlook Web Access			
Can you clean metadata on Lotus Domino Webmail/iNotes			
Can you clean documents going from BlackBerry devices			
Can you clean documents going from PDAs			
Can you clean documents on other mobile devices			
Can you perform a batch cleaning			
Can you clean a Zip archive			
Can you remove metadata from embedded emails			
Can you detect and remove metadata in password-protected documents			
Can you remove metadata without end-user interruption			
Are there enterprise-wide settings to automatically remove metadata from email attachments			
Can you preview cleaned attachments before emailing			
Can you export to an XML format			
Can you integrate into document repositories			
Matter Security			
Can you convert documents to PDF			
Can you enforce automatic conversion to PDF before emailing or transferring to removable media			

Matter Security (continued)			
Can you restrict sharing across ethical walls			
Can you enforce security when users are disconnected from the network			
Can you redact sensitive information from documents			
Can you classify documents to prevent unauthorized sharing			
Can you maintain an audit trail and risk report			

Osterman Research, Inc. does not provide legal advice. Nothing in this document constitutes legal advice, nor shall this document or any software product or other offering referenced herein serve as a substitute for the reader's compliance with any laws (including but not limited to any act, statue, regulation, rule, directive, administrative order, executive order, etc. (collectively, "Laws")) referenced in this document. If necessary, the reader should consult with competent legal counsel regarding any Laws referenced herein. Osterman Research, Inc. makes no representation or warranty regarding the completeness or accuracy of the information contained in this document.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND. ALL EXPRESS OR IMPLIED REPRESENTATIONS, CONDITIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE DETERMINED TO BE ILLEGAL.

About Workshare

Workshare is a leading provider of secure enterprise collaboration applications. Workshare allows individuals to easily create, share and manage high-value content anywhere, on any device. Workshare enhances the efficiency of the collaborative process by enabling content owners to accurately track and compare changes from contributors simultaneously. Workshare also reduces the commercial risk posed by inadvertently sharing confidential or sensitive documents. More than 1.8 million professionals in 70 countries use Workshare's award-winning desktop, mobile, tablet, and online applications. For more information visit www.workshare.com or follow Workshare on twitter at www.twitter. com/workshare.

Workshare empowers users with a document-centric collaboration experience, enabling them to review, make content-specific comments, and update documents based on user-defined permissions. This is managed with presence indicators, real-time alerts, and activity feeds. Workshare offers a range of integrations that enable customers to embed the power of Workshare collaboration and file sharing into existing enterprise content management and productivity applications such as Microsoft SharePoint and Microsoft Office (Excel, Word, and PowerPoint).

About Workshare Protect

With Workshare Protect you can remove hidden information (metadata) to protect against financial risk, a competitive disadvantage, or an embarrassing situation with costly consequences. Customizable with more than 25 document security options with unique settings for different types of recipients, Workshare Protect removes hidden information to reduce the risk of inadvertent exposure of confidential information with automated alerts and one-click removal of hidden information such as tracked changes, speaker notes, the author's name, and document editing time from Microsoft Office documents and PDF files.

About Workshare Point

Workshare Point enhances Microsoft SharePoint 2010 as a Document Management System by providing a document-centric front end that provides integration to both Microsoft Outlook and Office for better Microsoft SharePoint access and content management. Workshare Point integrates Microsoft SharePoint and Outlook. You can access SharePoint document libraries from the Outlook Navigation pane. A Reading Pane view in Outlook lets you perform tasks on documents within SharePoint. You can also easily drag-and-drop files from SharePoint into Outlook email either as links or attachments.

About Workshare Compare

Workshare Compare provides users with advanced document collaboration between multiple Word and PDF documents. The ability to compare images and Excel tables within documents, as well as the ability to accurately track document changes down to the paragraph, line, word, or character level, all add up to a streamlining of the entire document review process.

About Workshare Professional

Workshare Professional combines advanced document collaboration with the ability to keep documents secure. Users can expect accurate tracking even between the most complex documents while also benefitting from advanced security features when sharing documents, including the ability to remove potentially embarassing metadata or other information from any document. Additionally, Workshare Professional eliminates the need for multliple different document versions, maintaining all suggested document revisions within the original master document.

About Workshare Desktop

Workshare Desktop is ideal for individuals or organizations who want to collaborate, review, and comment on documents without the use of a web browser or Microsoft Office. Workshare Desktop also keeps local copies of files and folders in sync with the central cloud store, ensuring that everyone has the latest version and allowing users to continue working while offline.

About Workshare Mobile

Workshare Mobile makes collaboration and file sharing easy for users on the move. Users and their guests can view, share, and comment on files using any browser-enabled mobile device while all content remains auditable and protected from loss. Additionally, the Workshare Mobile App is available for all users wishing to connect their iPhone or iPad to the collaboration experience, seamlessly integrating Workshare into the iOS interface.