

Meeting Regulatory Challenges: Metadata in Court Submissions

A Workshare Report

Published April 2010



Executive Summary

Organizations have a critical need to remove metadata before sharing documents or submitting documents to the court. While organizations had been using electronic information in discovery actions for some time, the Federal Rules of Civil Procedure (FRCP) have elevated the importance of electronically stored information (ESI).

While metadata challenges have been reported in the mainstream media for years, a recent warning in February of 2010 from the U.S. District Court for the Western District of Pennsylvania highlights the continued challenges with educating the public on the dangers of metadata.

Information hidden or covered in a computer document can almost always be recovered. The way to avoid exposure is to ensure that sensitive information is not just visually hidden or made illegible, but is actually removed from the original document.

ABOUT THIS GUIDE

This guide discusses the challenges associated with cleaning metadata from documents and protecting organizations from the inadvertent disclosure of proprietary information. This guide also includes metadata removal best practices to ensure documents that are emailed or uploaded to court servers do not contain metadata.

Metadata Defined

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 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 the Properties feature.
- Documents can also 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.

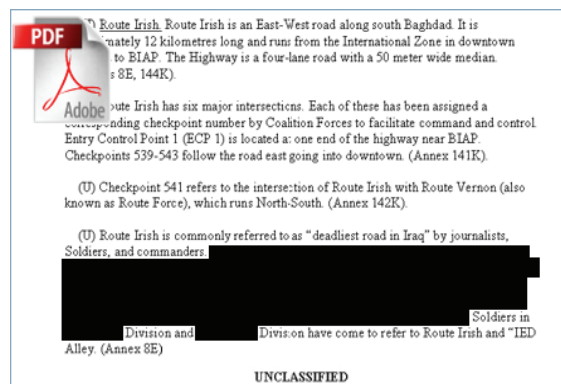
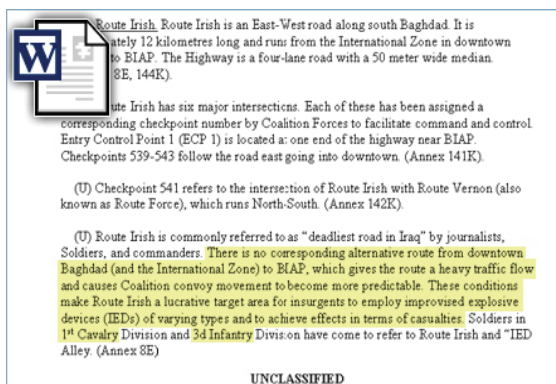
In short, metadata provides critical information about documents or files, but is rarely intended for display with the primary content included in the document.

Common Issues

The electronic document format introduces a few challenges and common mistakes when it comes to removing metadata.

- **Metadata and Document Properties:** In addition to the visible content of a document, most Microsoft Office documents contain hidden information. This information is often confidential and can lead to embarrassment for an organization, and to lost clients and lawsuits. Without a proper metadata removal application it can be quite challenging to remove all of the document metadata.
- **Redaction of Text and Images:** 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. While effective on printed materials, this technique does not work for electronic documents. Quite often the cover up can be removed to reveal the text underneath. Below is an example of a PDF with improperly redacted information.

The recipient of the Portable Document Format (PDF) version is able to copy and paste the information into Microsoft Word and view the text that was underneath the black graphic.



- **Comments and Tracked Changes:** A Microsoft Word user can also be at risk when converting a Word document to a PDF version. Quite often individuals will convert a Word document to PDF in order 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 a PDF file is created.

Meeting Regulatory Challenges

Electronic documents have been a key component of discovery actions for several years, but the amendments to the Federal Rules of Civil Procedure (FRCP) that took effect on December 1, 2006 significantly increased the risk for any organization that is 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 amendments essentially codified this practice. As a result, the metadata contained within electronic documents raises several important issues:

- **Metadata must be managed**

Metadata must be managed in a coordinated fashion throughout the lifecycle of all electronic documents. Organizations must establish policies about how metadata is created, managed, archived and deleted for all electronic document types.

- **Court submissions**

It is critical that the metadata in documents exchanged outside the organization is managed correctly. This includes documents being emailed and uploaded to network locations.

The occurrence of metadata in documents is common. In an effort to educate users, 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*

Metadata Removal Best Practices

Removing metadata before a document is converted or shared via email is essential to managing risk and confidentiality. It is essential to not only rely on technology tools to remove metadata, but to put the processes and training in place to ensure best practices are followed.

The following best practices should be followed to ensure your organization is protected from the risk of document metadata:

- Establish an enterprise-wide metadata policy and have a clear understanding of what metadata is and why it is a risk to the organization
- Review a report of existing metadata in documents before cleaning or sending to others
- Send or upload only those PDF files which have already been cleaned
- Remove all metadata from your template library to reduce the risk of any metadata being inherited from previously used documents
- Turn Track Changes off before creating a PDF
- Clean all Word documents before you create a PDF
- View the comments in your Microsoft Office application before sending documents to others to review if a history of comments exists

Workshare's Metadata Removal Solutions

Users of Workshare's metadata removal desktop applications have multiple options to remove metadata from documents.

Option 1: Clean metadata within Microsoft Word, Excel and PowerPoint

In order to clean metadata in your documents, perform these steps:

1. Open your Microsoft document
2. From the Workshare Panel or Menu, click Content Risk. For Office 2007, select the Workshare ribbon and click Content Risk. Workshare checks the document for content risk and displays the Content Risk page or Document Risk Report (in Excel and PowerPoint) or the Content Risk panel (in Word) that shows the details of the content risk found.
3. Click Remove. The Advanced Options dialog box is displayed.
4. Select the hidden data you want to remove by selecting the checkboxes to the left of the hidden data options.
5. After making your selections, click OK. The selected hidden data is removed from the document.
6. Save your document to keep the cleaned version.

Note: To clean document statistics, the document must be closed before cleaning. As a result, you cannot use the Content Risk method to clean document statistics. You must use one of the other cleaning options listed below. When the cleaned document is then opened, the following fields will have been reset:

- Date Created: Date and Time file is cleaned/sent
- Date Modified: Date and Time file is cleaned/sent
- Date Accessed: Always populated by Word as time the document is opened
- Last Saved By: Not Populated (Blank)
- Revision Number: 1
- Total Edit Time: Starts at 0 and updated by Word as time elapsed from the time the file was opened

Option 2: Clean email attachments

Every time you send an email with a Microsoft Office attachment, Workshare checks the attachment to see if it breaches any security policies. When you send an email that triggers a Clean action, Workshare notifies you that your Microsoft Office attachment(s) will be cleaned with the Workshare Protect dialog. If your administrator has enabled you to override the clean hidden data settings and you do not want to clean the attachment(s), you can select the Skip Cleaning checkbox in the Apply Action area.

1. After the Workshare Protect dialog is displayed, select the attachment in the Select File list and click the Go to Hidden Data Options hyperlink or select the Hidden Data Options tab. The Hidden Data Options tab displays the different hidden data cleaning options.
2. Select the hidden data that you want to remove by selecting or deselecting the relevant checkboxes.
3. Repeat these steps for additional attachments if required.
4. Click Send to send the email.

Workshare cleans the hidden data from the attached document(s) according to your settings before sending the email.

Option 3: Create a clean PDF document

To clean a document and convert it to PDF format, follow these steps:

1. Open your document in Microsoft Office 2003/XP and click PDF in the Workshare Panel or click the Convert to PDF button in the toolbar.
2. Open your document and then:
 - a. in Microsoft Word 2003/XP, select Convert to PDF from the Workshare menu.
 - b. In Microsoft Office 2007, click PDF in the Workshare Ribbon. The Create PDF dialog is displayed.
3. Click the PDF Security Options hyperlink. The PDF Security Options dialog is displayed. Select the Clean before PDF checkbox if you want to remove hidden data from the document before converting it to PDF.
4. Click Advanced Cleaning Options to specify which hidden data to remove. Select one or more of the security options individually or select the Risk Elements checkbox to select them all. Click Ok. On the PDF Security Options dialog, if required, set a password to protect the security.
5. Click OK.
6. If you want to create a PDF of part of the document only, select the Pages radio button and specify a page range. From the Print dropdown list, you can then select whether to PDF all pages in the specified range or only the odd or even pages.
7. If required, click Preview to view the document as a PDF.
8. Click OK. A Save As dialog is displayed. Give the PDF file a name, save it to a location of your choice, and click Save.

Option 4: Batch clean multiple documents

To clean multiple Microsoft Office documents at once, follow these steps:

1. From the Start menu, select All Programs, Workshare, and then Workshare Batch Clean. The Batch Clean dialog is displayed.
2. Select the documents you want to clean. Press Ctrl or the Shift key to select multiple documents.
3. Click Open. A second Batch Clean dialog is displayed. A complete list of hidden data that can be removed, reset or converted is listed in the dialog.
4. Select the hidden data you want to remove by selecting the checkboxes to the left of the hidden data options. Click Clean. All the selected files are cleaned using the same options.

Beyond The Desktop

With Workshare Protect Server organizations can protect proprietary information across their entire network with a server-based metadata removal solution. Workshare Protect Server secures confidential information being managed by Blackberry, iPhone and other mobile devices, removing metadata contained in Microsoft Word, Excel, PowerPoint, and PDF attachments. Workshare Protect Server also monitors attachments sent with Outlook Web access and Lotus iNotes.

Summary

Cleaning metadata from Microsoft Office documents, spreadsheets, and presentations is critical for every organization. To ensure that violations of legal, regulatory, or corporate policy are monitored in real time, organizations need to follow best practices and implement technology solutions to help manage their information and reduce their risks.

About Workshare

Workshare, a company with a twelve-year history providing metadata removal and document comparison technology, enables organizations to control and manage information securely. 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 14,000 organizations worldwide. For more information, visit www.workshare.com