

Why and How to Remove Metadata from Documents

Technology, Privacy, and eCommerce





In the Hybrid Counsel Trap article, our fictitious in-house team at Sunderland Manufacturing, deputy general counsel Barry Miles and his paralegal, Singita Patel, addressed risks to prevent attorney-client communications losing their privileged status. All technology communication problems solved,

right? Wrong!

Why is Metadata a Problem?

In a gender discrimination case venued in Illinois, Sunderland Manufacturing requests copies of all communications and other documents by the plaintiff, a former Sunderland employee, related to her employment. Jason Parks, Sunderland's external counsel, calls deputy general counsel Barry Miles to confirm receipt of the documents. He tells Barry that one of the documents, a PDF version of the plaintiff's electronic diary, contains dates when certain relevant disputed events occurred. When Jason "mined" the diary by reviewing its metadata, he discovered deleted entries in which the former employee described facts contradicting her claims which, if used, would materially undermine the plaintiff's case. She had also deleted her date of birth and social security number. Clearly, when she or her counsel converted the document to PDF, they did not intend the deleted information to be seen by Sunderland Manufacturing. Should Sunderland use the information?

Why Converting to PDF Isn't Enough

So how did Jason obtain the information in the first place? Why wasn't it enough for the plaintiff to edit the original Word document and convert it to PDF? When converting the format of a document, whether it is from word to PDF or another form, the document's original content is stored and preserved as metadata in the new PDF file. For example, if the plaintiff in this case redacted the electronic diary by placing a colored rectangle over the text, or using the highlighting tool in Word to highlight over the text with black color, the text would be preserved in the PDF. Covering the text does not erase it from the document; it merely forms a layer over the original text that can be easily removed with tools in Adobe® Acrobat® software.

Try this for yourself

1. Redact information as described above with your own sample document.
2. Convert the document into a PDF.
3. Save the document.
4. Depending on your PDF maker, use the "edit text and images" or "move or make changes to object" tool to move the paragraph where the highlighted text is. The word(s) you hid will appear as you drag the selected text away from the black highlighted area.
5. Now put the text back where it was.
6. Convert the PDF back into a Microsoft® Word® document. You will also be able to un-highlight the text and see the original words or information you intended to hide.

How Metadata can be viewed

In mining the PDF document, Jason Parks could have taken any of the following routes to find Plaintiff's hidden information:

A. If Plaintiff deleted the information by highlighting over the text, all Jason had to do to find the redacted information was either use the tools in Adobe to remove the text from behind the black rectangles, or convert the PDF back into a word document and un-highlight the redacted information.

B. If Plaintiff deleted portions of the document, Jason could either look at the basic properties of the document and find key information that was collected from the Word document during the conversion

— such as author, date of creation, date of modification, key words, or a former title. Additionally, depending on the version of software Plaintiff had, an entire original copy of the Word document could have been exported with the PDF and become accessible through the same properties function.

C. If Plaintiff had used “track changes” in the Word document, Jason could go into the PDF and — using the redaction tools Plaintiff should have used — seen the number of changes made, the substance of any track changes, or comments that were previously made in the document.

Ethical and Reporting Responsibilities

Whether or not a lawyer may mine for metadata, and whether that lawyer has a subsequent duty to disclose that mined information, depends on the jurisdiction. The Illinois Bar Association has not rendered a formal opinion on metadata ethics. Assuming the ABA model rules would apply in Jason’s case, under ABA Formal Opinions 06-442 and 05-437, Jason had both the *right* to look at the metadata and an ethical duty to *notify* the plaintiff regarding the information he saw.

Certain jurisdictions view metadata “mining” as ethically impermissible. The ABA does not share this view based on its interpretation of ABA Model Rule 4.4(b), which provides a sole requirement of providing notice to the sender of inadvertently sent information, as evidence of the drafters’ intent not to enumerate other specific restrictions on a receiving lawyer’s conduct. However, while the ABA does not require a receiving lawyer to “refrain from examining the materials,” the ABA does require that a lawyer receiving information or documents relating to their representation of a client, and know, or should know that the contents were inadvertently sent, to notify the sender.

Another component to consider is whether, based on the information found, Jason can move for sanctions against Plaintiff for improperly redacting the diary to exclude relevant information and dates. The most likely remedy will require Jason to request a discovery conference, allow for an in-camera inspection, and then allow the court to determine appropriate action. See Chinnici v. Central Dupage Hosp. Ass’n.

136 F.R.D. 464 (N.D. Ill. 1991) (Note: this case is accessible on Lexis, Westlaw, or in the Federal Reporter)

Steps to Remove Metadata

To avoid landing in the same quandary, both the in-house and external Sunderland legal teams routinely take the following steps to scrub documents before they are turned over to opposing parties:

1. Create a new copy of the document, and retitle it. This document will be used for redacting, the original remains on the office computer as a spare copy.
2. Physically delete any sensitive information. **NOTE:** employees should not re-size images or cover the text of these documents with black boxes or highlighting.
3. Create a new blank document. Select all of the text from the document you were working in, and then paste the information into the new blank document. Then, save that as a new document.
4. Convert the new document into a PDF. Before clicking convert, check the properties of the PDF conversion to ensure that no additional information is included in the PDF output.

Side Bar References

- The Proliferating Email Trap for Hybrid Counsel
- American Bar Association, Metadata Ethics Opinions around the U.S.
- ABA formal Opinion 05-437
- ABA formal Opinion 06-442
- Adobe Technical Note, Redaction of Confidential Information in Electronic Documents
- Fed. R. Civ. P. 5.2, Privacy Protection for Filings Made with the Court
- Microsoft® guidance on removing hidden data and personal information
- Chinnici v. Central Dupage Hosp. Ass'n, 136 F.R.D. 464 (N.D. Ill. 1991)

Adobe Acrobat is a registered trademark of Adobe Systems Incorporated in the United States and/or other countries.

Microsoft Word is a registered trademark of Microsoft Corporation in the United States and/or other countries

Spiwe L. Jefferson



Division General Counsel, Division General Counsel, Electrification Service

ABB

Spiwe Jefferson leads global legal and integrity functions across international operations for ABB's Electrification Service division. A sought-after speaker and thought leader on AI enterprise adoption, she has achieved over 710 hours (18 weeks) of AI-driven efficiency gains on one platform alone. Spiwe authors the *Mindful in 5* book series and podcast, providing leaders with actionable strategies for resilience and growth.

Jerry D. Hamilton



Managing Shareholder

Hamilton, Miller & Birthisel LLP

Jerry D. Hamilton is Managing Shareholder at Hamilton, Miller & Birthisel LLP and a Fellow of the Litigation Counsel of America.