

Cornell Law Library Scholarship@Cornell Law: A Digital Repository

Cornell Law School Graduate Student Papers


Cornell Law Student Papers

4-2014

"Out, Damned [Metadata]!"

Emily Shaw
Cornell Law School

Follow this and additional works at: http://scholarship.law.cornell.edu/lps_papers

 Part of the [Legal Profession Commons](#), and the [Science and Technology Commons](#)

Recommended Citation

Shaw, Emily, "'Out, Damned [Metadata]!'" (2014). *Cornell Law School Graduate Student Papers*. Paper 31.
http://scholarship.law.cornell.edu/lps_papers/31

This Article is brought to you for free and open access by the Cornell Law Student Papers at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law School Graduate Student Papers by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.

“Out, Damned [Metadata]!”¹

By Emily Shaw

I. Introduction

We live in exciting times; technology is evolving quickly. The legal profession, however, has a history of begrudging and delayed acceptance of new technology. Attorneys may be slow to learn new tricks, but when it comes to metadata, the usual reactionary behavior could be harmful to clients. It is imperative that attorneys understand the ethical and evidentiary issues that arise when metadata is disclosed, mishandled, discovered, or destroyed. This paper explores these issues and recommends best practices to avoid inadvertent disclosures and ethical violations. The structure of this paper is as follows: first, metadata is defined and explained. Second, I will explain potential harm that metadata can cause. Third, issues of confidentiality, attorney-client privilege will be explored. Fourth, I will explore some of the evidentiary concerns regarding discovery and destruction of metadata. Finally, the conclusion will recommend best practices for new and experienced attorneys to avoid metadata missteps and manage metadata with confidence.

II. Metadata Defined

Definitions vary with context and jurisdiction, but simply put, metadata is data about data. In the legal context, metadata is “all of the contextual, processing, and use information” associated with an electronic document.² John Kinas, director of information technology for the District of Columbia Bar, likened metadata to a price tag on a wedding gift—it’s a very useful piece of information when you’re buying the gift, but it becomes problematic if you accidentally leave it on.³ Metadata is useful and necessary for common computer applications. For example, the ability to ‘undo’ an edit in a Microsoft Word document relies on metadata that tracks the edit history of the document.⁴ Attorneys may be uncomfortable, to say the least, if their opposing counsel were able to click ‘undo’ to see the edit history of their client’s document. The metadata in a typical Microsoft Word document may include: the author’s name and initials, the name of the company or organization where the document was created, the names of previous document authors, the original text and any revisions, template information, digital comments, document versions, and hidden text.⁵

A competent attorney must understand how courts view metadata. The Southern District of New York defines metadata as “electronically-stored evidence that describes the ‘history, tracking, or management of an electronic document.’”⁶ The courts have recognized several distinct types of electronic metadata, including substantive metadata, system metadata, and embedded metadata.⁷ Substantive metadata (sometimes referred to as “application” metadata) is created as a function of the application software used to create the document or file.⁸ This includes information that instructs the computer how to properly display properties of a document such as the fonts, spacing, size, and color, as well as information that reflects modifications to the document, such as its edit history.⁹ This information transfers with the document because it is embedded in the file when it is moved or copied.¹⁰ A second category

of metadata is system metadata. System metadata “reflects information created by the user or by the organization’s information management system.”¹¹ This type of metadata is particularly helpful when attempting to search and sort large numbers of documents efficiently—both in the context of regular use and in e-discovery.¹² A third category of metadata is embedded metadata.¹³ Embedded metadata consists of “text, numbers, content, data, or other information that is directly or indirectly imputed into a [n]ative [f]ile by a user and which is not typically visible to the user viewing the output display” of the native file.¹⁴ This type of metadata includes spreadsheet formulas, hidden columns, hyperlinks, references, and database information.¹⁵ With this basic understanding of what metadata is, an attorney is prepared to confront issues that may arise with its intentional or unintentional disclosure or discovery.

III. Potential Pitfalls

Metadata is useful in document production. It was originally developed “by software programmers accustomed to working in collaborative environments where sharing information is commonplace.”¹⁶ Law firms also benefit from collaboration; in fact, Microsoft revealed in a 2001 press release that it solicited opinions from attorneys in developing Word 2002 because “the legal profession must have an efficient way to compare documents and incorporate text and formatting changes.”¹⁷ However, as much as the legal profession benefits from collaboration, it stands to lose a lot if that information falls into the wrong hands.

Examples of metadata mishaps range from embarrassing to catastrophic. In February 2003, British Prime Minister Tony Blair’s office published a dossier on Iraq’s security and intelligence organizations; this dossier was cited by Colin Powell in his address to the United Nations.¹⁸ This dossier was published as a Microsoft Word document.¹⁹ The metadata from the Word document revealed that, contrary to the government’s assertions, it had been drafted by civilians and that parts of it had been plagiarized from a thesis written more than ten years previously.²⁰

Portable Document Format (PDF) files have less metadata and are frequently regarded as safer from metadata mining than other file formats. While this is often the case, PDFs are not foolproof. Attorneys at Facebook learned this lesson the hard way. In 2009, large portions of the transcript of the settlement between Facebook and ConnectU were redacted and a PDF of the transcript was made publicly available.²¹ While the image of the PDF had blacked-out the redacted portions of the PDF, the metadata that provides the searchable text in a PDF was not altered.²² Thus, members of the Associated Press were able to simply copy and paste the sensitive information from the PDF.²³ The metadata revealed that Facebook’s internal valuation of the company was \$3.7 billion, \$8.88 per share.²⁴ This was far less than the \$15 billion valuation established by the Microsoft investment in 2007.²⁵

While disclosing settlement terms is highly embarrassing (and likely a breach of a confidentiality clause), active litigants have even more at stake. In 2004, the SCO group, an entity that licenses and sells Unix, filed a complaint in state court against DaimlerChrysler and AutoZone.²⁶ The metadata gleaned from the court-filed documents revealed that SCO group’s

attorneys had been building a case against Bank of America in federal court.²⁷ In fact, the Word document, when viewed under the “original showing markup” setting, identified Bank of America as a defendant until exactly 11:10 a.m. on February 18.²⁸

These examples have a few things in common. Each metadata leak was exposed by members of the press who had no duty not to disclose this publicly shared information. Each leak was found using rudimentary computer functions—Microsoft Word document settings and copy-paste commands. Each metadata leak caused a scandal. But most importantly, each metadata leak could have been prevented using basic metadata scrubbing procedures.

IV. Ethical Issues

The attorneys from the examples in section III clearly did not intend to leak confidential information through metadata, but these situations nevertheless raise various ethical issues. First, the disclosure may amount to a waiver of the attorney-client privilege. Second, the disclosure may amount to a waiver of work product protection. Third, while the above examples of metadata leaks were all exposed by third parties, it is important to note that opposing counsel that read inadvertently disclosed metadata may themselves be committing an ethical violation.

The attorney-client privilege is among the oldest and most fundamental protections in the American justice system.²⁹ This privilege protects communication between attorneys and clients in order to encourage full and frank communication between clients and their attorneys and thus “promote broader public interests in the observance of law and administration of justice.”³⁰ The Restatement (Third) of the Law Governing Lawyers defines the test of qualifying attorney-client communications as four elements: “(1) a communication (2) made between privileged persons (3) in confidence (4) for the purpose of obtaining or providing legal assistance for the client.”³¹ It is well-settled law that this privilege belongs solely to the client.³² However, courts acknowledge that “the attorney’s conduct may bind the client even in the absence of his express consent” if the attorney is acting under the authority granted to her as agent.³³ Implied waiver may occur when a client voluntarily discloses confidential communications to a non-essential third party.³⁴ Courts split into three distinct approaches when it comes to determining whether the attorney-client privilege has been waived through inadvertent disclosure.

The “Strict Approach” holds that inadvertent disclosure always waives the attorney-client privilege. This approach, sometimes called the “Wigmorean approach”, the “strict-liability approach”, or the “objective approach,” follows the teachings of Professor Wigmore, who believed that all disclosures “are not protected by the privilege, on the principle that, since the law has granted secrecy so far as its own process goes, it leaves it to the client and attorney to take measures of caution sufficient to prevent [disclosure]. The risk of insufficient precautions is upon the client.”³⁵ The D.C. Circuit is among the courts that follow this approach.³⁶ Proponents of this approach believe that a uniform application to the waiver will prevent abuse of the duty of confidentiality and encourage attorneys to take effective measures to prevent inadvertent

disclosures.³⁷ Conversely, critics say that this approach is unduly harsh and intrudes upon the attorney-client relationship because it may discourage clients from confiding in their attorney.³⁸ Furthermore, the Wigmorean view was devised in an age before liberal discovery practices and may be an impractical approach to today's high-volume document disclosures.

The "Lenient Approach" (also referred to as the intent-based approach) requires intent to disclose privileged material and fully protects from inadvertent disclosure.³⁹ Courts that follow this approach, including Federal courts in Florida and Illinois, reason that since waiver is often defined as "[t]he voluntary relinquishment or abandonment—express or implied—of a legal right or advantage,"⁴⁰ it is not possible to waive the privilege inadvertently.⁴¹ Proponents of this approach argue that this rule protects clients from their attorneys' negligence. This reflects the concept that the privilege belongs solely to the client. Like the Strict Approach, this approach also has the benefits of a uniform application and easy administration.⁴² However, critics of the Lenient Approach argue that this approach ignores the basic principles of agency law and leaves little incentive for attorneys to guard privileged information.⁴³

The third and most common approach, adopted by a majority of jurisdictions, is the Circumstances Approach—which, predictably, finds a middle ground between the Strict Approach and the Lenient Approach.⁴⁴ Courts that follow this approach examine all of the circumstances surrounding the inadvertent disclosure and allow waiver in only limited circumstances, such as when an attorney or client "fail[s] to take reasonable steps to maintain confidentiality."⁴⁵ Courts examine five factors to determine if the privilege is waived: "(1) the reasonableness of the precautions taken to prevent inadvertent disclosure in view of the extent of the document production; (2) the number of inadvertent disclosures; (3) the extent of the disclosure; (4) the promptness of measures taken to rectify the disclosure; and (5) whether the overriding interests of justice would or would not be served by relieving the party of its error."⁴⁶ Unsurprisingly, the gains in flexibility and fairness by adopting this approach cost some efficiency and predictability that go along with the uniform approaches. The bottom line for an attorney entrusted with confidential client information is to know your jurisdiction and act reasonably and competently.

The work product doctrine, codified in the Federal Rules of Civil Procedure 26(b)(3),⁴⁷ provides "qualified immunity for materials prepared in anticipation of litigation by a party, an attorney, or some other representative of a party."⁴⁸ When confronted with the issue of work product protection, nearly all jurisdictions follow a similar approach to the Circumstances Approach from the Attorney-Client Privilege jurisprudence.⁴⁹ Courts weigh the following five factors to determine whether a waiver has occurred: "(1) reasonableness of precautions taken to prevent disclosure, (2) time taken to rectify error, (3) scope of discovery, (4) extent of disclosure, and (5) overriding issues of fairness."⁵⁰

An attorney who receives inadvertently disclosed privileged information is faced with conflicting ethical obligations: they have both a duty to diligently represent their client and a duty to avoid dishonestly, fraud, deceit, or misrepresentation.⁵¹ This dilemma was addressed in

2002 by the Model Rule 4.4(b), which directs the receiving attorney to promptly notify the sender.⁵²

V. Evidentiary Issues

Metadata is an unsettled frontier in e-discovery and full treatment of the subject is well beyond the scope of this paper. Nonetheless, attorneys must understand the rules, or they risk sanctions for evidence spoliation or even criminal charges for destruction of evidence.⁵³ The watershed case on the subject, *Williams v. Sprint/United Management Co.*, held that a defendant who produces electronic files during discovery must also produce their corresponding metadata.⁵⁴ Three years after *Williams*, the Southern District of New York clarified some of the requirements for various types of metadata production in e-discovery.⁵⁵ Substantive metadata, which includes information such as prior edit history, editorial comments, and computer display instructions, “need not be routinely produced” unless the requesting party shows good cause.⁵⁶ System metadata, such as information about the author, date of creation, and date of modification are frequently considered irrelevant by courts.⁵⁷ System metadata may be relevant, however, if the authenticity of a document is questioned or dates of document creation are important to the case.⁵⁸ Embedded metadata, such as formulae in complicated spreadsheets, is “generally discoverable” and “should be produced as a matter of course.”⁵⁹ Importantly, the destruction of metadata can land a client and attorney in trouble for spoliation of evidence.⁶⁰

VI. Conclusion and Best Practices

As professionals entrusted with privileged client information and sophisticated data management responsibilities, it is imperative that new and experienced attorneys alike become familiar with the dangers of inadvertent metadata disclosure and evidence spoliation from the irresponsible destruction of metadata. This is a fine line that attorneys must walk. The most important thing to remember is that the moment a client is on notice of the pending litigation, metadata must be preserved with their corresponding electronic documents.⁶¹ At no point, however, should an attorney ever send a document with work product metadata. The best practice is to use metadata scrubbing software to ensure that outgoing documents and files that are sent to opposing counsel, third parties, or e-filed with a court are sent without potentially damaging metadata. Large firms typically have this kind of software integrated into their data management systems. Small firms and solo practitioners can purchase relatively inexpensive software programs that will remove metadata from outgoing documents and files. These software programs are extremely effective and are part of a responsible and reasonable effort to protect privileged information from inadvertent disclosure. In closing, all attorneys would be wise to heed John Kinas’s [and Lady Macbeth’s paraphrased] advice: “Scrub early, and scrub often.”⁶²

¹ Apologies and a happy 450th birthday to William Shakespeare.

² *Autotech Techs. Ltd. P’ship v. AutomationDirect.com, Inc.*, 248 F.R.D. 556, 557 n.1 (N.D. Ill. 2008).

³ Marilyn Caviccia, *How Clean is Your Document? What You Need to Know about Metadata*, 32 B. Leader 22 (2007-2008).

⁴ Adam Israel, *To Scrub or Not to Scrub: The Ethical Implications of Metadata and Electronic Data Creation, Exchange, and Discovery*, 60 Ala. L. Rev. 469, 472 (2009).

⁵ *Id.* at 472-73.

⁶ *Aguilar v. Immigration & Customs Enforcement Div. of the U.S. Dep't of Homeland Sec.*, 255 F.R.D. 350, 354 (S.D.N.Y. 2008).

⁷ *Id.* (citing United States District Court for the District of Maryland, *Suggested Protocol for Discovery of Electronically Stored Information* 25-28, <http://www.mdd.uscourts.gov/news/news/ESIProtocol.pdf>).

⁸ The Sedona Principles-Second Edition: Best Practices Recommendations and Principles for Addressing Electronic Document Production Cmt. 12a (Sedona Conference Working Group Series 2007), http://www.thesedonaconference.org/content/miscFiles/TSC_PRINCP_2nd_ed_607.pdf ("Sedona Principles 2d").

⁹ *Id.*

¹⁰ See *id.*

¹¹ *Id.*

¹² *Id.*

¹³ *Supra* note 6, at 354-355.

¹⁴ *Id.* (citing United States District Court for the District of Maryland, *Suggested Protocol for Discovery of Electronically Stored Information* 25-28, <http://www.mdd.uscourts.gov/news/news/ESIProtocol.pdf>).

¹⁵ *Id.*

¹⁶ *Supra* note 4, at 472.

¹⁷ *Id.*

¹⁸ Richard M. Smith, *Microsoft Word Bytes Tony Blair in the Butt*, <http://www.computerbytesman.com/privacy/blair.htm> (June 30, 2003) (last visited April 22, 2014).

¹⁹ *Id.*

²⁰ Arlen L. Tanner, *Metadata: Why the Fuss? A White Paper on Metadata*, 2 Bloomberg Law Reports-Technology Law 15 n. 20, available at <http://www.shb.com/attorneys/TannerArlen/MetadataWhytheFuss.pdf> (last visit April 22, 2014).

²¹ See <http://www.techcrunch.com/2009/02/11/the-ap-reveals-details-of-facebookconnectu-settlement-with-best-hack-ever/> (last visited April 20, 2014).

²² *Id.*

²³ *Id.*

²⁴ *Id.*

²⁵ *Id.*

²⁶ Stephen Shankland and Scott Ard, *Hidden text shows SCO prepped lawsuit against BofA*, http://news.cnet.com/2100-7344_3-5170073.html (March 4, 2004) (last visited April 20, 2014).

²⁷ *Id.*

²⁸ *Id.*

²⁹ See 8 JOHN HENRY WIGMORE, WIGMORE ON EVIDENCE § 2290, at 542 (1904).

³⁰ *Upjohn Co. v. United States*, 449 U.S. 383, 389 (1981).

³¹ RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 68 (2000).

³² See *In re von Bulow*, 828 F.2d 94, 100 (2d Cir. 1987).

³³ *Supra* note 4, at 472.

³⁴ Campbell Steele, *Attorneys Beware: Metadata's Impact on Privilege, Work Product, and the Ethical Rules*, 35 U. Mem. L. Rev. 911, 917 (2005).

³⁵ 8 WIGMORE, *supra* note 9, § 2325(3).

³⁶ *Underwater Storage, Inc. v. United States Rubber Co.*, 314 F. Supp. 546 (D.C. Cir. 1970).

³⁷ *Supra* note 34, at 917.

³⁸ *Id.*

³⁹ *Id.*

⁴⁰ BLACK'S LAW DICTIONARY 1611 (8th ed. 2004).

-
- ⁴¹ Supra note 4, at 477.
- ⁴² Supra note 34, at 921.
- ⁴³ Id.
- ⁴⁴ Id.
- ⁴⁵ Id. (citing *Bank Brussels Lambert v. Credit Lyonnaise (Suisse) S.A.*, 160 F.R.D. 437, 443 (S.D.N.Y. 1995)).
- ⁴⁶ *Edwards v. Whitaker*, 868 F. Supp. 226, 229 (M.D. Tenn. 1994).
- ⁴⁷ FED. R. CIV. P. 26(b)(3) (2004).
- ⁴⁸ Supra note (Steele), at 923-24 (citing CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, 2 FED. EVID. § 204 (2d. ed. 1994)).
- ⁴⁹ Surpa note 4, at 479.
- ⁵⁰ Id. (citing 6 JAMES WM. MOORE ET AL., MOORE'S FEDERAL PRACTICE § 26.70 [6] [c] (3d ed. 2008)).
- ⁵¹ Supra note 34. at 933.
- ⁵² MODEL RULES OF PROF'L CONDUCT R. 4.4(b).
- ⁵³ *McPeek v. Ashcroft*, 202 F.R.D. 31 (D.D.C. 2001).
- ⁵⁴ 230 F.R.D. 640 (D. Kan. 2005).
- ⁵⁵ Supra note 6, at 354.
- ⁵⁶ Id.
- ⁵⁷ See, e.g., *Mich. First Credit Union v. Cumis Ins. Soc'y, Inc.*, No. Civ. 05-74423, 2007 WL 4098213, at *2 (E.D.Mich. Nov.16, 2007).
- ⁵⁸ See *Hagenbuch v. 3B6 Sistemi Elettronici Industriali S.R.L.*, No. 04 Civ. 3109, 2006 WL 665005, at *3 (N.D.III. Mar.8, 2006).
- ⁵⁹ Supra note 6, at 355.
- ⁶⁰ See *McPeek v. Ashcroft*, 202 F.R.D. 31 (D.D.C. 2001).
- ⁶¹ See id.
- ⁶² Marilyn Caviccia, *How Clean is Your Document? What You Need to Know about Metadata*, 32 B. Leader 22 (2007-2008).