I Never Meta Data I Didn't Like

By Judge Herbert B. Dixon Jr.

The unoriginal title of this column most likely expresses a feeling that is the opposite of those experienced by judges, lawyers, and regular people the first time they confront issues involving metadata in electronically stored information (ESI). Disputes involving paper documents have never been fun. Metadata makes the enjoyment quotient of electronic documents even lower.

What Is Metadata?

Metadata is often described as "data about data" or "hidden data." Commentators have called these descriptions unhelpful. It is more helpful to think of metadata as information about electronically stored files that is hidden within the files themselves or in a linked database.[1]

In some circumstances, both the existence and the purpose of metadata are obvious. Consider, for example, that metadata underlies the "undo edit" function within your word processing program; it allows you to undo all the changes you made since you last saved your document. Another example of metadata can be found in the various formulae encoded within a spreadsheet that define the mathematical relationship between cells. Metadata provides information about the time and date a document was created, the computer and network on which it was created, the number of times the document was revised, previous revisions made (including deletions and additions), contributing authors, the electronic format of the document, hidden text, deleted comments, dates the document was accessed, and more. In sum, metadata is the DNA of the electronic world.[2]

There are two main types of metadata: application metadata and system metadata. Application metadata is embedded data that permits someone having the requisite knowledge and skill to determine a document's secrets, i.e., information not explicitly stated on the face of it. This kind of information may include, e.g., the author's name, revisions, and so forth. System metadata is not embedded within a document. Rather, it is information on the computer's hard drive or within the computer system's memory that details the size of every electronic file or document, its location within the computer system, and additional information about the file's secrets, including some of the same information to be found within the application metadata. As an aside, computer forensics experts can confirm document spoliation (i.e., tampering) when they note differences between the application metadata and system metadata related to a specific electronic document.

Metadata is important for a number of reasons. For starters, it provides information about the storage location of each electronic file. Today's desktop and laptop computers can easily store four to nine million pages of documents, and they can often store, not only simple text files, but also images that require far greater amounts of storage. Without trying to be too precise, think of this amount of storage as the equivalent of two library floors of unlabeled, randomly placed law books. Without metadata indicating the location of each book, you would have to pull each book in the library to determine if that is the one you are looking for.

Metadata ESI Discovery Issues

An issue that often arises in ESI discovery disputes is whether the requesting party is entitled to receive intact or embedded metadata. This process starts when the requesting party asks for documents in "native" format, i.e., in the same electronic format the document was first created in and maintained. Those who have been initiated in this process recognize that documents produced in native format come with metadata—meaning detailed information about the original document itself and its subsequent history. In the context of legal proceedings, the potential importance of metadata is readily apparent, as metadata can provide information about the creation and alteration dates of an electronic document, the dates it was accessed or read, its author, hidden text, deleted comments, other changes in the document made after it was created, and more.

As federal case law has developed, a party is generally entitled to the production of metadata when documents in native format are sought in the initial document request and the producing party has not yet produced the documents in any form.[3] That's when some interesting issues arise, including those in instances

Judge Herbert B. Dixon Jr. is the technology columnist for *The Judges' Journal*. He sits on the Superior Court of the District of Columbia and is a former chair of the National Conference of State Trial Judges. He is chair of the Judicial Division's Court Technology Committee and a member of the Planning Board for ABA TECHSHOW. He can be reached at Herbert.Dixon@dcsc.gov.

where proprietary software only available from the producing party is needed to read, search, and organize the native documents, and the producing party objects to sharing its proprietary business information . . . but those types of issues are beyond the scope of this article.

Metadata Embarrassments

Below are a few high-profile embarrassments that occurred seemingly because the people involved suffered from an apparent knowledge deficiency about metadata.

2004

SCO Group, Inc. sent warning letters to more than 1,500 of the world's largest companies threatening a lawsuit for damages in the event those companies failed to obtain a license from SCO for use of Unix and Linux software. Later, carrying out its planned litigation strategy, SCO sued IBM, seeking damages of $5 billion, and in separate lawsuits also sued others, including DaimlerChrysler. The embarrassment to SCO was that metadata in the electronic copy of the complaint against DaimlerChrysler revealed dates and times of other versions of the complaint and deleted information, including that the Bank of America was the named defendant in an earlier version of the complaint, a deleted comment inquiring whether Bank of America "received one of the SCO letters sent to the Fortune 1500," a change of the basis of liability from copyright infringement and violations of the Digital Millennium Copyright Act to violations of a Unix software agreement with SCO, and changes in the particular relief requested.[4] Taken together, all of these changes revealed that the grounds for the lawsuit were, in actuality, only shifting sands. In 2007, following a series of unsuccessful lawsuits concerning Unix and Linux software, SCO filed a petition for Chapter 11 bankruptcy protection.

2005

The Pentagon completed a classified report of its investigation into the shooting of an Italian secret agent in Iraq by a U.S. soldier, and it prepared an unclassified public version of the report in PDF format for posting on the Internet. The unclassified version of the report was created by electronically blacking out substantial portions of the classified sections of the report. The blacked-out, classified details included unpleasant and embarrassing information about the occurrence and names of others who were involved or present, as well as sensitive information concerning military procedure at roadblocks. To the embarrassment of the Pentagon and Italian authorities, knowledgeable persons with appropriate software who downloaded the public document were able to read the full report by merely copying and pasting the contents of the PDF document into a Word document.[5] So much for military intelligence.

2006

Justice Department prosecutors filed court documents concerning their grand jury investigation into steroids used by professional athletes. Regrettably, the prosecutors made nearly the identical mistake committed by Pentagon officials the year before. They electronically blacked-out portions of the PDF documents to redact confidential grand jury evidence from public view. Once again, knowledgeable persons with appropriate software copied the document contents into a word processing program and were able to read the details of some of the confidential grand jury evidence.[6]

> Metadata can provide information about deleted comments and other changes.

2007

Someone at the Defense Intelligence Agency posted on the agency's website a PowerPoint presentation that included a slide depicting the trend of dollar awards for intelligence projects. Unfortunately, the presentation was posted in native format, which allowed anyone downloading the file to take a peek at a closely guarded national secret—the size of the intelligence budget overseen by the Director of National Intelligence. The presentation was removed immediately after the agency learned of press reports about the error.[7]

Also in 2007, Federal Trade Commission (FTC) lawyers filed court documents concerning the Commission's investigation of the anticompetitive effects of a takeover of Wild Oats Markets by Whole Foods Markets. The FTC lawyers ineffectively redacted the documents; the result was that the public gained access to information that Whole Foods considered confidential and proprietary business information.[8]

2008

A default judgment in the amount of $5,247,781.45 was entered as a sanction for defendants' failure to comply with discovery orders, along with an award of $645,760 in attorneys' fees and costs. In addition to the defendants' failure to produce financial documents, computer forensics experts' study of metadata in one of the defendants' computers revealed that financial documents relevant to the litigation had been deleted and a software program that included a data-wiping utility had been used several times to overwrite data and certain deleted files, which made it impossible to identify everything that might have been erased.[9]

The Common Themes

Metadata embarrassments often occur when the metadata accompanying an electronic document reveals confidential information. These embarrassments also occur when the metadata reveals information contradicting the public position of an individual circulating the document, shows that a document was plagiarized or taken from a questionable source, or contains unflattering comments. Imagine the resulting fallout from the release of a document containing metadata that candidly revealed a judge's or lawyer's concerns about the correctness of a position taken in the document itself. Not a pretty sight.

Metadata 101: How to View It

Suppose you are in academia and have recently received a reviewer's report concerning the article that you submitted for publication. As you know, these peer review critiques are supposed to be anonymous. However, you might easily be able to learn information about the person who wrote that stinging review. Microsoft Word automatically tags every document with information taken from the computer on which the document was prepared, including the author's name, company's name, and even statistical information such as the number of revisions and total editing time. To possibly learn the identity of your tormentor, click the "File" menu, then click "Properties," and then click the "Summary" tab. Voila! If the reviewer or the publication staff members have not taken adequate precautions to delete those tags, you may learn the name of your anonymous critic, and more.[10]

Metadata Ethics for Lawyers

In ongoing litigation, generally, responding lawyers have an ethical obligation to preserve metadata in electronic documents produced in response to discovery requests, and receiving lawyers have carte blanche to study metadata in all documents received. Conversely, lawyers exchanging documents that are not the subject of ongoing litigation face the potential minefield created by the differences in states' approaches to the ethics of metadata. Uniformly, though, the ethics rules require lawyers to take proper precautions to prevent inadvertent disclosure of confidential or privileged information when transmitting paper or electronic documents. However, aside from the rules regarding discovery of ESI in ongoing litigation, states' rules diverge with respect to metadata mining, i.e., the act of deliberately seeking out and viewing metadata in documents received. In some states, mining is permitted and in others, it is prohibited. And in still others, the receiving lawyer is prohibited from reviewing metadata sent by an opponent only when the lawyer has knowledge that the metadata was inadvertently sent. Under the latter rule, the receiving lawyer must consult with the sending lawyer to determine whether the metadata includes work product or confidential or privileged information.[11]

Metadata Removal

Where there is no legal obligation to preserve a document's metadata, the most straightforward way to avoid embarrassment is to use computer software to remove the hidden data. This kind of software is called a metadata remover or metadata stripper. Newly issued versions of some applications today come with a feature to remove hidden data. Absent a preinstalled feature to remove metadata, stand-alone software of this type is available both commercially and as freeware. Just do a Google search for "metadata removal" and choose from the many selections that you'll find. The brief time needed to remove metadata is worth the effort to avoid the embarrassment of your recipient finding metadata in the document that reveals indecisive vacillation or unflattering comments contained in earlier drafts but deleted from the final version.

Conclusion

If metadata is not removed from an electronic document before it is distributed, the sender may share with everyone who acquires the document information that is confidential, privileged, or highly embarrassing.

For attorneys producing requested discovery, a significant concern is the unwitting production of privileged information that may lead to waiver of the privilege. For judges issuing a decision in a contested case or lawyers communicating a settlement demand or offer, there should always be a concern about whether the metadata might reveal details of the initial draft and previous versions of the document.

Take heed, everyone: when talking about metadata, what you can't see can hurt you. Be careful of the written documents that you create on your computer. What's needed now is a metadata analogy for the old saying that "even a fish wouldn't get into trouble if he kept his mouth shut." ■

Endnotes

1. Institute for the Advancement of the American Legal System, Navigating the Hazards of E-Discovery: A Manual for Judges in State Courts Across the Nation, (University of Denver, 2007).

2. Craig Ball, *Beyond Data About Data: The Litigator's Guide to Metadata*, 2005, *available at* http://www.craigball.com/metadata.pdf.

3. Aguilar et al. v. Immigration & Customs Enforcement Div., 2008 U.S. Dist. LEXIS 97018, 2008 WL 5062700 (S.D.N.Y. Nov. 20, 2008).

4. Stephen Shankland & Scott Ard, *Hidden Text Shows SCO Prepped Lawsuit Against BofA*, CNET News, http://news.cnet.com/2100-7344_3-5170073.html.

5. Kieren McCarthy, *That Classified U.S. Military Report's Secrets In Full*, The Register, *available at* http://www.theregister.co.uk/2005/05/03/military_report_secrets/.

6. Josh Gerstein, *Prosecutors Demand Reporters Testify in Steroid Case*, N.Y. Sun, June 22, 2006, *available* at http://www.nysun.com/national/prosecutors-demand-reporters-testify-in-steroid/34920.

7. Ralph Losey, *Metadata Mistake by Top Spy Agency*, e-Discovery Team Blog, June 12, 2007, http://ralphlosey.wordpress.com/2007/06/12/metadata-mistake-by-top-spy-agency/.

8. Associated Press, *Documents Describe Whole Foods' Strategy*, N.Y. Times, *available at* http://www.nytimes.com/2007/08/15/business/15food.htm?_r=2.

9. Southern New England Tel. Co. v. Global Naps, Inc., 251 F.R.D. 82 (D. Conn 2008) (Second Amended Ruling).

10. Jeffrey R. Young, *Microsoft Word's Hidden Tags Reveal Once-Anonymous Peer Reviewers*, Chron. Higher Educ., Apr. 21, 2006, *available at* http://chronicle.com/free/v52/i33/33a04101.htm.

11. Donna Payne, *The New Metadata Rules*, Oct. 2008, *available at* http://www.iupui.edu/~facinfo/I350/new%20metadata%20rules.pdf.