**Forensic Implications of Metadata in Electronic Files**

*By John Ruhnka and John W. Bagby*

## Understanding Metadata

In this digital age most business activities are transacted and recorded using networked information systems. Business and accounting records are prepared, reviewed, audited, and preserved in electronic form, called "electronically stored information" or ESI. It is estimated that 94% to 99% of all business records are created and maintained in electronic form (*The National Law Journal,* July 17, 2006, p. S1) and most are never transformed into hard copy. A unique characteristic of electronic records is that they include hidden "metadata" that contains extensive information about the creation of a file, including "MAC dates" (the dates a file was modified, accessed and created), the date last printed, and if deleted, when it was deleted and by whom. Metadata also reveals the file name and the path where a document is located on a computer or network, identifies the computer on which it was created, the name of the person who last saved the document, the number of revisions made, and any document ID or client/matter properties added to the document. Emails contain metadata that indicates the sender's address book information, the date a message was sent, received, replied to, forwarded, and to whom copies were sent, and the existence of attachments. Metadata has been called "the electronic equivalent of DNA" and has the capability to shed light on the origins, context, authenticity, and distribution of electronic evidence (Craig Ball, *Beyond Data about Data: The Litigator's Guide to Metadata* 2005). This article provides an overview for accountants of how metadata is generated by software applications they use, and the potential significance of metadata in the electronic business records and emails that they produce and work with. This is important both in accounting forensics and in litigation support activities.

## Metadata Production, Location and Access

Two types of metadata are associated with correspondence, spreadsheets and other electronic files used by accountants. *Application metadata* is automatically created by applications software and is embedded in every file created or edited using that software. Operating systems that control individual computers, servers and communications systems create *systems metadata,* which assigns file allocation table fields (file name, creation, length, and use) to all files stored on the system so that the operating system can identify and locate

that file for future use. Systems metadata resides in the system registry of the computer system or server used to access and store that file.

Many CPA's use Microsoft Office programs including Word, Excel, PowerPoint, and Outlook (for emails and electronic calendars). All of these applications automatically produce dozens of fields (types) of application metadata for each file created using these programs. Application and systems metadata fields are created and updated for MS Word, Excel, and PowerPoint files each time a file is created, opened, or used, as well as optional "track changes" or "create versions" information that a user may intentionally create in a file. Adobe Acrobat software creates detailed metadata path information which can provide forensic information on files created on or stored on these programs as well,

**Significance of Metadata in Litigation**

A 2007 survey of litigation activities of 253 U.S. corporations revealed that 83% of respondents had new lawsuits filed against them in 2006 (Fulbright & Jaworski, *Fourth Annual Litigation Trends Survey Findings*, October 2007). The most common subject of these lawsuits was labor/employment, contract enforcement and personal injuries. Litigation was also significant at smaller companies surveyed, 17% had at least one lawsuit claiming $20 million or more, and at mid-sized companies, 98% reported one or more lawsuits of $20 million or larger. After a law suit is filed, a pre-trial "discovery" phase occurs in both state and federal courts during which the litigants are required to identify and disclose ("produce") all information in their possession that is requested by the opposition, which is potentially relevant to the subject matter of the litigation. Since most settlements in litigation occur before a trial is held, electronic records and emails disclosed and evaluated by the parties during the discovery phase can often be determinative of litigation outcomes.

Once a lawsuit is filed or a party has been served with a document preservation request, a legal "litigation hold" arises which requires litigants to preserve all evidence under their control potentially relevant to the subject matter of the litigation. In some circumstances a legal duty to preserve potentially relevant evidence can even arise <u>before</u> a lawsuit is filed. The watershed 2003 *Zubulake* discovery ruling imposed legal duties to preserve potentially relevant evidence as soon as litigation is "reasonably anticipated" (*Zubulake v UBS Warburg,* 2003 WL 22410619 at *4, S.D.N.Y. 2003).

Because metadata in electronic files reveals forensic information about the creation, authorship, history and even intent of a document, it can play a potentially critical role in

litigation outcomes. In the Merck & Co. *Vioxx* product liability litigation which resulted in a large judgment against Merck, *The New England Journal of Medicine* reported that residual "tracked changes" accidently left in a Merck internal document indicated that Merck knew of potential dangerous side effects of Vioxx including heart attacks two years *before* placing the drug on the market (*Forbes*, Dec. 8, 2005). The general rule on disclosing the metadata associated with files demanded in litigation has been stated as follows: "when a party is ordered to produce electronic documents as they are maintained in the ordinary course of business, the producing party *should produce the electronic documents with their metadata intact*, unless that party timely objects to production of metadata, the parties agree that the metadata need not be produced, or the producing party requests a protective order." (*Williams v. Sprint/United Mgmt Co.*, 230 F.R.D. 640, D. Kan. Sept. 29, 2005).

**Preserving Metadata when Reviewing or Producing Files**

Larger organizations are often involved in frequent litigation which requires the entity to identify, preserve and disclose potentially relevant electronic files in response to successive legal discovery demands. To effectively manage this complicated process which can have significant liability implications, it is sound practice to institute an enterprise-wide "ESI Discovery Team" to manage and coordinate this complex and tremendously costly discovery process. An ESI Discovery Team includes key decision makers who need to be involved in the on-going process of planning for and responding to discovery requests. This typically includes the CIO, IT system managers, in-house legal counsel, representatives from administrative units most closely involved in the litigation (e.g.,HR director in a wrongful termination lawsuit), and outside litigation counsel and third party legal and electronic discovery consultants and forensic experts if they will be involved. The ESI Discovery Team designs an organization's "litigation hold" procedures and deploys "litigation holds" often involving multiple and overlapping litigation, over all enterprise locations.

An entity subject to a litigation hold must act quickly to prepare a written "preservation plan" that identifies all information potentially relevant to the litigation across all enterprise locations. The identification process can use key words describing the subject matter of the litigation, identify specific users whose emails, instant messaging and voice mails may be relevant, and notify identified users to preserve all data on desktop PCs, laptops and messaging devices. The 2006 Federal Rules of Civil Procedure (FRCP) require both accessible and "inaccessible" ESI, such as network and server back-up tapes, to be preserved so over-writing

or reuse of back-up tapes that may include emails potentially relevant to the litigation must be immediately halted. The 2006 FRCP Rule 26(f) requires a "meet-and-confer" conference to occur early in the litigation among the parties to negotiate the scope of discovery by and for each side - what files are to be collected, reviewed for potential relevance and, if relevant, produced - as well as the format in which files are to be produced and whether they will include metadata, and the timetable for discovery.

The FRCP contains a default preference for delivery of electronic files in "native file format" (the format in which the data is ordinarily preserved) *including all associated metadata*. Thus potentially relevant ESI needs to be preserved in native file formats with metadata intact before the multiple steps involved in collection and review of files for relevance are initiated. Opening a file for review alters its metadata and could be viewed as "tampering" with evidence. Parties producing files requested for discovery need to be able to show an unbroken "chain of custody" to assure admissibility as evidence and to avoid judicial sanctions for altering or tampering with document files. To ensure this, it is advisable to make a secure "snapshot" digital record of all potentially-relevant enterprise server systems and files which is separately archived before any review for potential relevance is conducted so that original files and metadata remain intact. If litigation parties disagree about the format in which files are to be produced or whether file and system metadata is to be included, the federal courts will look at the potential relevance of metadata to the issues in dispute. In an options backdating case, for example, metadata showing the dates of successive entries contained in options documents could be critical. A second consideration is the cost of producing file metadata. If metadata already exists in the native file formats it is more likely to be required, whereas it is less likely to be required if it is not present in the native file format and must be reconstituted from other sources.

**Forensic Uses of Metadata**

Forensics accounting provides an evidentiary basis for economic transactions and reporting events by identifying the process of capturing, using, storing and transmitting business and financial data in an enterprise information system. This can involve manual processes such as data entry, computations, verifications, and communications with others, in conjunction with an enterprise's IT and network systems. Metadata can help to identify the human and system actions in information systems and can be used to investigate and verify fraud, abuse, mistakes, or system failures, and can help to establish elements such as causation, timing and extent of knowledge or *mens rea* (guilty knowledge) which are at issue in criminal or civil litigation. An

example of the forensic use of metadata is in stock options backdating investigations, where the integrity of the dates entered facially on written option documents is disputed. In *Ryan v. Gifford, (*2007 Del. Ch. LEXIS 168) the Delaware Chancery Court ordered respondents to produce the disputed stock option documents in an electronic format that would permit examination of all metadata associated with the documents, noting that "Maxim's special committee as well as Deloitte & Touche undoubtedly reviewed metadata as part of their investigation into the backdating problems at Maxim."

In *Williams v. Sprint/United Management Co*., 230 F.R.D. 640 (D. Kan. 2005) the plaintiff in an age discrimination lawsuit sought an Excel workbook in its native file format. However, defendant Sprint stripped out all metadata from the Excel spreadsheet files that it produced, arguing that the metadata could reveal privileged information which Sprint had a right to withhold from production (the formulas and calculations used to derive information in the Excel spreadsheets which were linked to spreadsheet cells). The Federal Court held that *blanket* withholding of metadata from the requested accounting records went too far, and ordered Sprint to produce all of the metadata in its accounting files as maintained in the ordinary course of business, except for specific metadata which it claimed was protected by attorney-client privilege

**Confidentiality & Malpractice Implications – Metadata in Client Files**

Some potentially relevant information demanded in litigation, including attorney-client communications and litigation work product including associated metadata, may be withheld from disclosure as "privileged" information, subject to judicial review and approval of such claims of privilege. Claims of privilege must be identified in a "privilege log" which identifies all withheld files as to author, recipients, subject matter, and dates. The privilege log alerts opposing parties to the fact that potentially relevant information has been withheld and the identity of participants and dates enables opponents to review and challenge claims they believed to be unjustified. The 2006 FRCP amendments impose a faster pace for discovery which increases the risk of accidental disclosure of privileged documents, but also provides that parties may request a "claw back" (court-ordered return) of privileged files or trade secrets in the event of inadvertent disclosure. Litigants may also request court protective orders that prohibit the disclosure of proprietary, confidential, or private information that has been accessed by an adversary or its experts.

CPA firms play an important role in providing electronic discovery and forensic services in litigation. The 2007 Socha-Gelbmann Electronic Discovery survey (http://www.sochaconsulting.com/2007survey.htm) indicates that $2.6 b was spent on electronic discovery services in 2006, and that CPA firms are increasingly significant vendors in this arena (Ernst & Young and KPMG were ranked in the top ten electronic discovery service providers in 2007). CPA's who provide forensic information and damage calculations for clients need to be aware of the liability implications of metadata contained in client files. Inadvertent disclosure of client file metadata could result in a waiver of subsequent client claims of legal privilege for the metadata, or enable opponents to use metadata against clients' interests. CPA's are increasingly being held to the same malpractice standard as lawyers. *Mattco-Forge, Inc. v. Arthur Young & Co. (*6 Cal.Rptr.2d 281 Cal. Ct. App. 1992) involved a suit by Mattco against CPA firm, Arthur Young, hired as an expert witness and damages consultant to assist them in a lawsuit against GE, Mattco claimed that Arthur Young negligently provided unsubstantiated calculations for the profits allegedly lost because General Electric had struck Mattco from its supplier list. Because original estimate sheets were not available for all contracts, Arthur Young had asked Mattco to prepare noncontemporaneous estimates for the missing estimate sheets. These estimates, not identified as being noncontemporaneous, were turned over to GE who used them to have Mattco's legal claims dismissed. The California Appellate Court noted that in today's technologically driven litigation, engineers, physicians, real estate appraisers and many other professionals, including accountants, hired to assist a party in preparing and presenting a legal case can play as great a role in the organization, shaping and evaluation of their clients case as do lawyers and accordingly should be held to the same malpractice standards.

**Managing Enterprise Metadata**

While metadata should not, without prior judicial approval, be intentionally altered or removed from documents subject to a litigation hold or demanded in litigation, metadata may be removed in the ordinary course of business from both enterprise and client documents as necessary to preserve enterprise and client confidentiality as well as to safeguard proprietary information. AICPA Code of Professional conduct, Rule 301 *Confidential Client Information* provides "A member in public practice shall not disclose any confidential client information without the specific consent of the client." If a CPA who was assisting a client with bid calculations were to send an amended version of a bid proposal to the opposing side including metadata that revealed that the client had initially approved much higher bid amounts, the CPA

could be liable for breach of client confidentiality or even a malpractice claim for jeopardizing the contract. Potential liability for disclosure of metadata harmful to client interests means that metadata confidentiality policies that will pass muster with both legal discovery rules and AICPA ethics rules, including pre-release metadata viewing and "scrubbing" (intentional metadata removal) of security-sensitive files should conducted on an enterprise-wide basis and not be left up to individual discretion. CPA firm personnel must learn the necessary technical skills to both view and to remove metadata from electronic files,

Metadata is viewable in several ways. Basic metadata in MS Office documents is viewable from the "file ‘'properties" option, under tabs labeled as "General", "Statistics, and "Contents". Word will reveal to any user a Word document's authors, the date of creation, last modification, number of revisions, and where the document is stored. If optional used-added features such as "track changes" or "comments" were enabled when a Word document was created or edited, selecting the "Insert" option on the toolbar menu and selecting "comments" will reveal who made specific edits to a document and when. There are also commercially-available metadata viewers which can be used to access a much larger array of metadata in a file. For examples, see h*ttp://www.payneconsulting.com/products* and *http://www.docscrubber.com/.* Payne Group also produces Metadata Assistant, a widely-used metadata "scrubber". Detailed instructions on removing metadata from electronic files are beyond the scope of this article. Microsoft offers an MS Office 2003/XP add-in called "Remove Hidden Data" which can remove most but not all metadata from MS Office 2003 documents and Microsoft also offers "Office Document Inspector" for MS Office 2007 which can remove most metadata from Word, Excel and PowerPoint files. For a white paper on the uses and limitations of MS *Office Document Inspector* see *http://esqinc.com/Content/WhitePapers/Document-Inspector.php.*

(Sidebar) **Enterprise Metadata Security**

- Establish enterprise-wide (not individualized or optional) security policies to identify and tag metadata- sensitive projects and work product that could harm firm or client interests if metadata is disclosed. Require security verification of metadata content of security-tagged electronic work product prior to release.

- Train personnel in awareness of liability for disclosure of metadata – legal discovery implications; preserving client confidentiality and "privileged" information status; potential malpractice for impairing client or firm interests.
- All personnel should know how to view and to remove metadata in electronic files; enterprise-wide technical advice and assistance should be available.
- Utilize an enterprise-wide “ESI Discovery Team” to direct and monitor all identification, screening and production of ESI in response to litigation discovery requests.
- Be able to advise clients on the forensic implications of metadata and associated legal and evidentiary exposure - e.g. backdating options, bid/proposal revisions.

---

***John Ruhnka, JD, LLM**, is the Bard Family Term Professor of Entrepreneurship at the business school of the University of Colorado Denver, Denver, Colo. **John W. Bagby, JD**, is a professor and co director of the Institute for Information Policy in the college of information sciences and technology at the Pennsylvania State University, University Park, Pa.*