ARMA INTERNATIONAL
EDUCATIONAL
FOUNDATION
RESEARCH · EDUCATION · SCHOLARSHIP

Metadata in Court:

What RIM, Legal and IT Need to Know

By John Isaza, Esq.
November 2010

Research Project Underwritten by:

The ARMA International Educational Foundation Endowment

1609 Terrie Drive

Pittsburgh PA 15241 USA

www.armaedfoundation.org

Table of Contents

INTRODUCTION

When it comes to declaring electronic records, information governance professionals have been struggling for some time with the issue of what metadata[1] to preserve. This may not be as much of an issue with native files,[2] so long as they are preserved intact upon creation, or when it comes to data stored in various and sometimes sophisticated platforms such as back-up or archival systems, and other enterprise systems that may automatically store all metadata in native format. This is, in fact, the preferred mode of production currently advocated by some counsel and courts. Unfortunately, when it comes to preservation of documents generated by common business applications (often unstructured data found in both large and small organizations[3]) for records management purposes, the question becomes more difficult. What metadata fields are considered usable or relevant in determining what to capture at the end of the day if a record is or has been migrated to another format?[4]

Similarly, when it comes to discovery[5], organizations struggle with the scope of how much metadata to preserve for pending and anticipated litigation or agency investigations.[6] In 2004 and in 2007, the ARMA International Educational Foundation published two articles dealing with the thorny issue of spoliation, discussed in Section II(B) below, and the related preservation obligations that pertain to electronically stored information ("ESI").[7] The 2007 article, in particular, addressed the issue of scope of production in the context of what information to put on hold when faced with pending or reasonably anticipated litigation. As of December 2006, the Federal Rules of Civil Procedure (the "FRCP") had been amended to reflect growing technology and the need to address ESI. Although not specifically mentioned in the revised rules, when it comes to ESI a necessary question arises about how these rules should apply to metadata. How then does an organization avoid spoliation sanctions, considering how relatively easy it is to alter metadata during the day to day operation of any business application?

This paper seeks to apply case law and legal think tank[8] opinions that have touched upon the issue of metadata to glean best practices on what metadata to preserve in litigation as compared to metadata that should be preserved for records management purposes. Regrettably, the courts have never directly addressed preservation of metadata

---

[1] See Appendix for explanation of Metadata.

[2] See Appendix for explanation of Native Files.

[3] See Appendix for explanation of Unstructured Data.

[4] See Appendix for explanation of Migration.

[5] See Appendix for explanation of Discovery.

[6] Another key concern for counsel in particular is the inadvertent disclosure of privileged information that may be contained in metadata. The complexity of privilege issues is not addressed here, as this topic is not within the scope and purpose of this paper.

[7] Isaza, John. "Anticipated Litigation": New Case Developments to Determine Triggering Events & Scope of Production." (October 2007) and "Legal Holds & Spoliation: Identifying a Checklist of Considerations that Trigger the Duty to Preserve." (October 2004), available at www.armaedfoundation.org .

[8] Legal think tanks include groups of lawyers, judges, professors and other practitioners who assemble to address best practices for a myriad of legal issues, including electronic discovery.

for pure records management purposes. And, to the extent they have addressed metadata, they have limited treatment of the issue to only a handful of business applications, mostly desktop. They are otherwise relatively silent when it comes to enterprise-wide systems such as enterprise resource planning systems (ERP)[9] and the records that may be generated from those systems.[10]

To arrive at an informed analysis, this paper begins by briefly exploring the role of metadata in authentication of records. This analysis will help the reader put in perspective how metadata ultimately could affect the admissibility[11] of a record or document in court, and thus lead to an understanding of what and why certain metadata is critical. Next the paper will explore the general concepts of metadata and spoliation, including the leading legal think tank opinions on the issue. The paper then will address recent concrete examples where the courts have ruled on the discoverability of metadata. Finally, the paper will conclude with a list of discernible patterns of preservation requirements for information governance professionals to glean in setting policies and procedures regarding the capture of metadata for both records management and e-discovery preservation. Given the limited treatment of the issue in courts, this paper will focus on the typical varieties of applications[12] that yield documents or records where metadata is most often sought or litigated. These include email (and attachments), word processing documents, spreadsheets, presentation documents (e.g. Power Point), graphics, animations, images, audio, video and public records. From discussion of these samples, the reader may be able to glean metadata fields to preserve for more customized applications or even enterprise-wide systems that facilitate record-keeping.

I. ROLE OF METADATA IN AUTHENTICATION

Authentication[13] of records is one of the most critical reasons why metadata merits discussion. The outcome of several cases has turned on whether sufficient metadata existed to authenticate key documents at issue.

In *Lorraine v. Markel*, for instance, the plaintiff yacht owners sought to enforce a private arbitrator's award in their favor against an insurance company for damages sustained to their yacht.[14] Addressing the admissibility of ESI, the court noted that

> "computerized data ... raise unique issues concerning accuracy and authenticity ... The integrity of data may be compromised in the

---

[9] See Appendix for explanation of ERP systems.

[10] This could be because these systems tend to have metadata preservation capabilities that may preserve the metadata with the native file, or at least facilitate metadata preservation.

[11] The issue of admissibility (i.e., whether a document is actually admitted in court) is not explored here. Admissibility is first and foremost fact specific, and it entails one of the most complex areas of evidentiary law.

[12] See Appendix for explanation of Applications.

[13] See Appendix for definition of Authentication.

[14] 241 FRD 534 (D. Md. May 4, 2007)

> course of discovery by improper search and retrieval techniques, data conversion, or mishandling."[15]

"Whenever ESI is offered into evidence," the court explains, "this data must survive the rules of evidence. Even if relevant…, the ESI also must be authentic as required by rule 901(a)…"[16] Quoting rule 901(a), the court further details what constitutes authentication:

> "the requirement of authentication or identification as a condition precedent to admissibility is satisfied by evidence sufficient to support a finding that the matter in question is what its proponent claims."[17]

Various Sections of Rule 901 of the Federal Rules of Evidence are used to authenticate ESI. Rule 901(b)(1), for example, allows for authentication by "testimony that a matter is what it is claimed to be." This is the type of evidence most often used to authenticate emails. Because the "knowledge" required for a valid testimony is construed broadly, courts have held "a witness with personal knowledge," such as one who can identify an email, may be enough to authenticate emails.

The more recently drafted Rule 901(b)(9) further provides that "[e]vidence describing a process or system used to produce a result and showing that the process or system produces an accurate result" can be used to authenticate records.[18] Similarly, under Rule 901(b)(4) (calling for consideration of "distinctive characteristics and the like… [including] appearance, contents, substance, internal patterns, or other distinctive characteristics, taken in conjunction with circumstances.") electronic evidence can be authenticated by examining the metadata for the evidence.[19] If a party requests production of ESI in its "native format," for example, the metadata included in this format would show the "date, time and identity of the creator of an electronic record, as well changes made to it."[20] While not foolproof, the "metadata is a distinctive characteristic of all electronic evidence that can be used to authenticate it." [21]

FRE Rule 901 is thus a key factor used by the courts to authenticate electronic records. Rule 901(b)(4) has been singled out to compel the examination of metadata as a key factor in the process. The *Lorraine* case also highlights the importance of requesting ESI in its native format, so that it captures key metadata such as date, time and identity of the creator of an electronic record, and any changes made to the record. Such information will make the record easier to authenticate and more likely to be admissible in court.

---

[15] *Id*. at 544 (citing *In re Vee Vinhee*, 2005 Bankr. LEXIS 2602 (9th Cir. Bankr. Panel Dec. 16, 2005).
[16] *Id*. at 538.
[17] *Id*. at 542.
[18] *Id*. at 537.
[19] *Id.*
[20] *Id*. at 548.
[21] *Id.*

II. METADATA, SPOLIATION & THE THINK TANKS

Given the role of metadata in the authentication of records, a more detailed analysis of metadata is warranted from the court's point of view. This includes leading think tank treatment of the issue, in addition to the concept of spoliation as it could apply to metadata.

A) Types of Metadata

Metadata describes the history, tracking, or management of an electronic document. It includes hidden text, formatting codes, formulae, and other information associated with an electronic document. The Southern District of New York in *Aguilar v. Immigration & Customs Enforcement Division* identified three types of metadata: substantive, system, and embedded.[22] These distinctions illustrate the multiple layers of metadata, although admittedly courts seem to have done very little with these distinctions.

1. Substantive Metadata

Substantive metadata, also known as application metadata, is created as a function of the application software used to create the document or file. It reflects the substantive changes made by the user.[23] Examples of this type of metadata are found in documents that track modifications, "such as prior edits or editorial comments" and includes "data that instructs the computer how to display the fonts and spacing in a document."[24] Substantive metadata is "embedded in the document it describes and remains with the document when it is moved or copied." Substantive metadata could be relevant in litigation, because it shows what changes were made to a document, even if the document does not reflect these changes on its face. In other words, substantive metadata can be accessed even when no longer visible on the face of the document.

From a records and information management ("RIM") point of view, substantive metadata (i.e., reflective of substantive changes to a document) is most commensurate with drafts of a record. The history of changes to a document (other than its creator or author) does not serve much use once the ultimate record is declared. However, for discovery purposes, substantive metadata could be relevant. Thus, substantive metadata may not be necessary to declare and classify a record, but it should be preserved once litigation is pending or threatened.

---

[22] *Aguilar v. Immigration & Customs Enforcement Div*., 255 F.R.D. 350, 354 (S.D.N.Y. 2008), *citing The Sedona Principles, Second Edition: Best Practices Recommendations and Principles for Addressing Electronic Document Production*, Cmt. 12a (Sedona Conference Working Group Series 2007),

[23] *Id*. at 354, *citing* United States District Court for the District of Maryland; *Suggested Protocol for Discovery of Electronically Stored. Information* 25-28, http://www.mdd.uscourts.govinews/news/ESIProtocol.pdf ("Maryland Protocol") downloaded August 18, 2010;
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*"), downloaded August 18, 2010.

[24] *Id*.

2. System Metadata

System metadata "reflects information created by the user or by the organization's information management system." [25] Examples of system metadata include "file names and extensions, sizes, creation dates and latest modification dates."[26] Information regarding a digitally stored photograph, for example, would be considered system metadata.[27] System metadata is relevant if the authenticity of a document is questioned. It is also relevant if establishing who received what information and when are important to the claims or defenses of a party. This type of metadata makes electronic documents more functional. It significantly improves a party's ability to access, search, and sort large numbers of documents efficiently.

From a RIM point of view system metadata (e.g., file names and extensions, sizes, creation dates) is most relevant. This data could be crucial to authenticating the record for any purpose where authenticity of the record is paramount, like archival or historical uses. For discovery purposes, system metadata would be equally relevant.

3. Embedded Metadata

Embedded metadata consists of "text, numbers, content, data, or other information" that appears in native files with spreadsheet formulas, hidden columns, hyperlinks, references and fields, and database information. [28] Some examples of embedded metadata include spreadsheet formulas, hidden columns, externally or internally linked files (such as sound files), hyperlinks, references and fields, and database information. Embedded metadata is "critical to understanding a database application."[29] For instance, a complicated spreadsheet may be difficult to comprehend without the ability to view the formulas underlying the output in each cell.

Embedded metadata is thus relevant from both a RIM point of view and litigation. After all, the court in *Aguilar,* discussed later in this paper, stressed that embedded metadata is crucial to understanding a document or database application.

Despite the three distinctions noted for types of metadata, the courts ultimately decide what metadata must be produced on a case by case basis. Specifically what the courts have considered essential in fact-based scenarios must be addressed. Some examples may glean a pattern of metadata that courts uniformly deem relevant, discussed in Section III below. Failure to preserve may otherwise subject the organization to claims of spoliation, which is discussed next.

---

[25] *Id.*

[26] *Matter of Irwin v. Onondaga County Resource Recovery Agency*, 72, A.D.3d 314, 2010 NY Slip Op 1238, 6 (N.Y. App. Div. 4th Dep't 2010).

[27] *Id.*

[28] *Id.* at 13.

[29] *Aguilar, supra,* at 354.

B) Spoliation

One of the greatest dangers created by the failure to effectively manage records, especially during threatened or pending litigation, is spoliation. Spoliation is legally defined as "the destruction or significant alteration of evidence, or the failure to preserve property for another's use as evidence in pending or reasonably foreseeable litigation."[30] The Sedona Conference, discussed next, also defines it as "the destruction of records or properties, such as metadata**,** that may be relevant to ongoing or anticipated litigation, government investigation or audit."[31] Penalties for spoliation range from financial sanctions to rulings against admissibility of key documents at trial, adverse inference instructions to the jury (i.e., where a jury may be instructed by the judge to assume that any missing information would have been harmful to the non-producing party), dismissal of the case or, in extreme circumstances, jail.[32] Given the dynamic nature of metadata, it is especially easy to alter it and thus commit spoliation.

In *Krumwiede v. Brighton Assoc.,* a breach of contract case, spoliation of the metadata itself was not at issue, but it was used to establish spoliation. The court considered spoliation sanctions on whether plaintiff intentionally destroyed evidence by deleting files relevant to the case.[33] The plaintiff's laptop contained file entries listing critical metadata such as the last time certain files were accessed and deleted - actions that took place after plaintiff initiated the court proceedings. The court found that "[d]espite [his] duty to preserve evidence," the plaintiff's actions constituted "clear and convincing evidence" of willful destruction.[34] The court explained the plaintiff's suspicious activities, notably a "spike in activity" which involved destroying files as evidenced by the metadata that revealed a spike in activity. Such destruction was sufficient to show spoliation. [35] The court concluded there was "clear and convincing evidence" establishing plaintiff acted "willfully and in bad faith when he continued to alter, modify, and destroy evidence" despite his duty to preserve such evidence.[36]

Although no Federal or state court has issued an opinion on sanctions for spoliation strictly of metadata, it follows from *Krumwiede* and even the Sedona Conference's definition of spoliation that destruction of metadata is indeed subject to potentially severe sanctions.

---

[30] *Zubulake v. UBS Warburg*, 220 F.R.D. 212, 217 (S.D.N.Y. 2004). For more information on spoliation, including remedies, please see Isaza, John. "Legal Holds & Spoliation: Identifying a Checklist of Considerations that Trigger the Duty to Preserve." (October 2004).

[31] E-Discovery & Digital Information Management 48 (2d ed. 2007), *available at* http://www.thesedonaconference.org/content/miscFiles/TSCGlossary_12_07.pdf ("SEDONA CONFERENCE GLOSSARY")

[32] For a more complete list of sanctions, *see supra* Isaza, cited in footnote 28; *see also, Victor Stanley v. Creative Pipe, Inc.*, 2010 U.S. Dist. LEXIS 93644 (D. Maryland 2010), where the court recommended two years of jail time as a sanction for spoliation.

[33] *Krumwiede v. Brighton Assocs*, 2006 U.S. Dist. LEXIS 31669 (N.D. Illinois 2006).

[34] *Id.*

[35] *Id.*

[36] *Id.* at 26.

C) Leading Think Tank Treatment of Metadata

The bench has grown increasingly aware over the last couple of years that ESI discovery can be used effectively as a weapon in litigation to bring parties to their knees. Accordingly, courts have begun to show greater disdain for ESI discovery battles. Arguments over production of metadata are thus very low on a court's priority list, absent a showing of relevance on the part of the requesting party. Despite the court's animosity towards battles over ESI and its metadata, it is difficult indeed to set a steadfast rule to eliminate production of metadata. There is at least one advocate for establishing a rule that preservation of metadata should not be expected or relevant.[37] In fact, The Sedona Conference originally proposed a presumption against production of metadata. The below noted think tanks, however, compel a contrary argument that metadata should be considered at all times.[38]

The Sedona Conference

From the outset, the *Sedona Guidelines: Best Practice Guidelines and Commentary for Managing Information & Records in the Electronic Age* (the "Sedona Guidelines") state at Guideline 3f that "absent a legal requirement to the contrary, organizations are *not* required to preserve metadata; but may find it useful to do so in some instances."[39] That said, Guideline 3f goes on to state the following:

> "Because metadata may provide a wealth of information that can allow an organization to better retain and organize its information, organizations may find the retention and use of metadata to be beneficial from an organizational or operational perspective. Furthermore, organizations may wish to consider retaining sufficient metadata about records to ensure the trustworthiness of the records for organizational, fiscal, legal and historical purposes. And many organizations employ information and records management programs that specifically use metadata tags to cull and organize information. Finally, it may be that certain metadata is critical to an organization's ability to audit and track access to information so that it can, for example, identify and stop any improper access to sensitive information by unauthorized personnel. Thus, for some organizations it may be unworkable and unwise to routinely discard metadata. An organization should consider the best format in which to retain information to meet good business practices as well as legal requirements."[40]

---

[37] Joseph, Gregory, "Electronic Discovery and Other Problems," paper presented at the annual ABA Conference, Section of Science & Technology, August 5-7, 2010.

[38] The 2010 Civil Litigation Conference at Duke Law School, which took place on August 30, 2010, did not publish an official report as of the date this paper was published. Its discussions on metadata, if any, are not covered in this paper.

[39] *The Sedona Guidelines, 2d: Best Practices Recommendations and Principles for Addressing Electronic Document Production,* Guideline 3f at 28 (Sedona Conference Working Group Series November 2007) (emphasis added).

[40] *Id.*

The second edition of the Sedona Principles' *Best Practices Recommendations and Principles for Addressing Electronic Document Production* (the "Sedona Principles") further notes the need to preserve and produce metadata in discovery. Removed is the presumption against production of metadata suggested by the first edition. The pertinent section, Principle 12, was revised to read, "…production should be made in the form or forms in which the information is ordinarily maintained or in a reasonably usable form, *taking into account the need to produce reasonably accessible metadata…"*[41] The commentary to Principle 12 was divided into four subsections, a through d, that further elucidated the criteria for determining when metadata should be produced and how it should be protected. Comments 12.a and 12.b are most instructive. Comment 12.a. recognizes that production should vary according to the facts of a specific case.[42] Specifically, "[p]arties and counsel should consider: (a) what metadata is ordinarily maintained; (b) the potential relevance of the metadata to the dispute and (c) the importance of reasonably accessible metadata to facilitating the parties' review, production, and use of the information."[43]

Comment 12.b. considers the formats used for collection and production and the distinction between "ordinarily maintained" versus "reasonably usable."[44] The Comment highlights how FRCP Rule 34(b) "provides that in the absence of agreement or a specific court order, a producing party should produce electronically stored information either in the form in which it is 'ordinarily maintained' or in a 'reasonably usable' form."[45] To the extent Rule 34(b) calls for information to be produced in "reasonably usable form," courts have recognized this to mean that searchability of the document should not be limited.[46]

As far as the phrase "ordinarily maintained," it is not necessarily the same thing as the form in which the metadata was created.[47] "There are occasions," for example, "when business considerations involve the migration or transfer of electronically stored information to other applications or systems."[48] The comment further elucidates this example: "customer information may routinely be gathered by an organization from Internet-based forms, then collected in a relational database for further business use."[49] Similarly, what is reasonably usable depends on the particular circumstances of a case. Some factors counsel should consider include:

---

[41] *Aguilar, supra.,* at 356 (Emphasis added in court opinion).
[42] *The Sedona Principles, 2d: Best Practices Recommendations and Principles for Addressing Electronic Document Production* Cmt. 12a at 60 (Sedona Conference Working Group Series June 2007).
[43] *Id.* at 61.
[44] *Id.*
[45] *Id*. at 62.
[46] *Aguilar, supra.,* at 358, *citing The Sedona Principles.*
[47] *Id.*
[48] *Id.*
[49] *Id.*

> "(a) the forms most likely to provide the information needed to establish the relevant facts of the case; (b) the need for metadata to organize and search the information produced; (c) whether the information sought is reasonably accessible in the forms requested; and (d) the requesting party's own ability to effectively manage and use the information in the forms requested."[50]

The Sedona Principles's newest version of Principle 12 reflects the growing need to recognize metadata as an important aspect of discovery in a case. Not only is metadata deemed important, but the Conference recognizes the need for metadata to be produced in a "reasonably" accessible format to ensure it can be used properly and with minimal expenditure of resources. Due consideration should be given to the facts of a specific case and by extension to the form in which records are ordinarily maintained.[51]

The *2010 Sedona Commentary on Legal Holds* also reiterates the importance of Principle 12 as it relates to metadata. This Commentary states that "[i]t is often advisable to maintain sources of ESI in their native formats with metadata to preserve the ability to make production in some variant of a native file format, if necessary."[52] Thus, the 2010 Commentary further drives home the stated preference for retaining documents, at least for litigation purposes, in their native format.

Given the somewhat conflicting back drop of the Sedona Conference, organizations should have procedures or, at minimum, guidelines that set what metadata to retain for records management purposes. This will be the organization's first line of defense when questioned regarding record retention practices and ultimately regarding preservation for discovery purposes. Of course, once litigation is threatened or pending, organizations should strive to preserve all metadata that is, at minimum, reasonably accessible, if not preferably in native format. Once discussions commence with opposing counsel, further details on precisely what metadata to preserve may come to light.

The Seventh Circuit Electronic Discovery Pilot Program

The Seventh Circuit's Electronic Discovery Pilot Program (the "Seventh Circuit Pilot Program") was formed to streamline the arduous process of electronic discovery. It makes a valiant effort at drawing a clear line for metadata that does NOT need to be preserved, albeit a very thin line. The Seventh Circuit's Pilot Program's Report on Phase One, Principle 2.04 (Scope of Preservation), states that "data in metadata fields that are frequently updated, such as last opened dates" are "not

---

[50] *Id*.
[51] *The Sedona Principles, 2d: Best Practices Recommendations and Principles for Addressing Electronic Document Production* Cmt. 12a at 60 (Sedona Conference Working Group Series 2007), http://www.thesedonaconference.org/content/miscFiles/TSC_PRINCP_2nd_ed_607.pdf, downloaded August 14, 2010.
[52] *The Sedona Conference Journal*, Volume 11, page 277 (Fall 2010), downloaded from http://www.thesedonaconference.org/dltForm?did=legal_holds_sept_2010.pdf, October 31, 2010.

discoverable in most cases…"[53] The committee's reasoning for this principle is based on the fact that "many litigants do not have ESI collection tools that can collect data without affecting such metadata fields… Because the last-opened metadata field rarely will be the key to resolving most civil cases, the increased cost generally will not be warranted." Despite this reasoning, the Seventh Circuit Pilot Program goes on to say that:

> "These categories are not placed beyond the scope of discovery in all cases. The purpose of this Principle is simply to require litigants to promptly notify their adversary if they believe their case necessitates preservation and production of ESI in one or more of these categories. However, in raising the preservation of these categories, the demanding party should keep in mind that vague and overly broad preservation demands and responses are discouraged in Principle 2.03."

The Seventh Circuit Pilot Program addresses the cost of producing metadata in litigation. If nothing else, it attempts to carve out a clear line of demarcation for a type of metadata that should be presumptively beyond the scope of discovery - data in metadata fields that are frequently updated, such as last opened dates. In essence, certain categories of metadata should be beyond the scope of discovery simply because of the high costs associated with it, coupled with their comparative lack of probative value (i.e., is the metadata worth all the effort for the document at issue).

The Maryland Protocol

In response to the 2006 amendments to FRCP Sections 16, 26, 33, 34, 37, and 45, and Form 35, the United States District Court for the District of Maryland was one of the first, if not the first, to release a suggested Protocol for Discovery of Electronically Stored Information (the "Maryland Protocol").[54] In addition to addressing issues such as the importance of a discovery conference and the scope of discovery, the Maryland Protocol discusses Rule 26(b)(2)(C) cost-benefit factors. Even if metadata is relevant, the Maryland Protocol states, it is still subject to a cost-benefit shifting under the FRCP.[55] The Maryland Protocol expands upon the notions recognized in the Seventh Circuit Program discussed above, namely that costs are something that must be considered when evaluating metadata, but not to the extent that otherwise relevant evidence should be excluded in part because of the high cost of its capture and production.

As with the Seventh Circuit Pilot Program, the Maryland Protocol does little to inform what metadata to preserve, focusing instead on the high costs associated with its production and the resulting cost-shifting analysis.

---

[53] http://www.insd.uscourts.gov/News/7thphase%20one.pdf, downloaded August 28, 2010.
[54] http://www.mdd.uscourts.gov/news/news/ESIProtocol.pdf, downloaded August 28, 2010.
[55] *Id*. at 25.

The Judge's Pocket Guide

Published in 2007, the Judge's Pocket Guide (the "Guide") was "designed to help federal judges manage the discovery of electronically stored information (ESI)."[56] The Guide puts a different spin on the cost issue. It seems to neutralize the edicts of the Seventh Circuit Pilot Program and the Maryland Protocol by reminding judges that cost should NOT be the only factor. Specifically, the Guide recognizes the dynamic nature of ESI, and the implications of these differences for discovery. "The dynamic nature of ESI," for example, "makes it vital that a data producer institute 'litigation holds' to preserve information that may be discoverable, often even before the lawsuit is filed." Included in the questions the Guide seeks to answer is how to determine when the evidence produced has been stripped of metadata.[57] The Guide explains such information may be important regardless of the high cost and effort required to produce it:

> "… because deleted or backup information may be available, parties may request its production, even though restoring, retrieving, and producing it may require expensive and burdensome computer forensic work that is out of proportion to the reasonable discovery needs of the requesting party."[58]

The Guide recognizes that ESI, including metadata, should be managed before lawsuits are filed. The Guide logically leads to a distinction between preservation of data versus production. Preserving ESI, and as much metadata as is reasonably possible before formal discovery begins, allows parties to have realistic expectations of what to expect later in production, including the effort necessary for that production.

III. COURT RULINGS ON EXAMPLES OF METADATA

Though a general discussion about the role of metadata to authenticate records and think tank opinions serve to illustrate general guidelines and principles, specific case examples dealing with different kinds of ESI and the required metadata fields inform the issue more specifically. Regrettably, few types of ESI and their related metadata have been analyzed in court decisions. However, the examples below serve as a starting point to begin identifying some discernible patterns.

A) The Leading Metadata Case

The leading case regarding metadata is arguably *Aguilar v. Immigration and Customs*.[59] The court's in-depth treatment of the issue merits a brief discussion of the facts and its legal analysis of the issues.

---

[56] *Managing Discovery of Electronic Information: A Pocket Guide for Judges*, http://www.fjc.gov/public/pdf.nsf/lookup/eldscpkt.pdf/$file/eldscpkt.pdf, downloaded August 30, 2010

[57] *Id*. at 14.

[58] *Id*. at 3-4.

[59] *Aguilar, supra*, 255 F.R.D. 350.

The Facts in the *Aguilar* Case

In *Aguilar*, thirty Latino plaintiffs brought a civil rights class action suit against the Immigration and Customs Enforcement Division of the United States Department of Homeland Security ("ICE").[60] The plaintiffs maintained the ICE and its employees "subjected them to unlawful searches of their homes in violation of the Fourth Amendment."[61] The parties in Aguilar failed to discuss metadata before requesting production of documents. Accordingly, the plaintiffs failed to mention metadata until the defendants "had almost completed their document collection efforts."[62] Several months into the start of litigation, the parties discussed metadata during a conference call, where the plaintiff requested emails and electronic documents produced in their Tagged Imaged File Format (or "TIFF") "with a corresponding load file containing metadata fields and extracted text" as well as spreadsheets and databases in their native format.[63] The defendants objected to production of the data requested in TIFF form, instead proposing the data be produced in PDF form. With regards to the metadata, the defendants agreed to provide it "if the plaintiffs were able to demonstrate that the metadata associated with a particular document was relevant to their claims." [64]

Summary of Aguilar from a Discovery Point of View

The court noted in *Aguilar* that while the FRCP do not directly address metadata, it is still "subject to the general rules of discovery."[65] The court considered also the Sedona Principles, based on the Sedona Conference and its published articles regarding ESI.

Addressing existing case law, the *Aguilar* court recognized that other courts "generally have ordered the production of metadata when it is sought in the initial document request and the producing party has not yet produced the documents in any form."[66] Conversely, where such a request is made after the documents are produced in another form, courts have denied requests, deeming the metadata irrelevant. The court reasoned that if the requested format was relevant, it would have been requested sooner. The court also addressed the applicability of FRCP Rule 34's reference to producing the information in "reasonably usable form"[67] and stated other courts had recognized this to mean that searchability of the document should not be limited.[68]

Further addressing the importance of metadata, the court cited *Williams v. Sprint*, stating, "while metadata may add little to one's comprehension of a word processing

---

[60] *Id.* at 352.
[61] *Id.*
[62] *Id*. at 353.
[63] *Id*.
[64] *Id.*
[65] *Id*. at 355.
[66] *Id.* at 357.
[67] *Id.* at 358.
[68] *Id*.; *see also United Cent. Bank v. Kanan Fashions, Inc.*, 2010 U.S. Dist. LEXIS 83700 (N.D. Illinois 2010). where the court challenged the defendants to show "a particularized need for metadata of certain documents" before ruling on whether the metadata needed to be produced by the plaintiff bank.

document, it is often critical to understanding a database application."[69] Determining the importance of metadata requires an analysis of whether such data is necessary to an understanding of the document's application. In the case of a spreadsheet, for example, the need for metadata "depends upon the complexity and purpose of the spreadsheet."[70]

Summary of Aguilar from a RIM Point of View

The *Aguilar* case initially focuses on the need to produce metadata when the parties have requested the information from each other. However, it is clear from the subtext that, at minimum, a record needs to be preserved in a "reasonably usable form." This would, of course, include related metadata to help deem the record "usable." *Aguilar* cites Principle 12 of the Sedona Principles in treating the metadata that should be ordinarily maintained, relevance, and importance of the metadata to the parties' review, production, and use of the information. This treatment is of little use to the records manager. Relevance, for instance, cannot be determined until litigation becomes threatened or pending. Ultimately, then, usability of the record becomes the key factor for determining what metadata to preserve from a RIM point of view.

B) Emails

The *Aguilar* case provides a useful example of metadata issues that can arise in emails. The court explains that one of the consequences of producing an email in one form instead of another is the possibility of not knowing who was blind copied (or bcc'd) on an email or in which folders the emails were saved.[71] The plaintiffs in the *Aguilar* case argued this information was relevant because it supported their theory that the ICE searches were part of a pattern of unconstitutional conduct. Specifically, the plaintiffs would know who was bcc'd if they could see the metadata, which in turn would reveal who knew of the emails and took part in the alleged wrongful conduct. Additionally, the underlying metadata would have made searching the emails easier. The defendants were ultimately ordered to produce "any emails … received with metadata attached in a form that contains that metadata."[72]

C) Word Processing Documents and PowerPoint Presentations

In *Aguilar*, the plaintiffs sought metadata regarding the following for all Word, Excel, and PowerPoint documents:

- Date created
- Date modified
- Modified by

The plaintiffs claimed this information was relevant because without it they could not properly search the documents or ascertain "'who knew what when."[73] Pointing to the

---

[69] *Id*. at 354, citing, *Williams v. Sprint* , 230 FRD 640 (D. Kansas 2005).
[70] *Id*.
[71] *Id*. at 359.
[72] *Id*. at 364.
[73] *Id*. at 361.

fact that the documents had already been produced in PDF, the court held the metadata was unnecessary, even though the documents were not so numerous as to make a search burdensome. The court also explained the plaintiff failed to show why knowing the who and when would have any bearing on whether their constitutional rights were infringed. Again, addressing relevance, the court reasoned that if the metadata was essential to the plaintiff's claims, they would have sought it initially instead of waiting. Although the court did not grant the plaintiff's request here, it is clear the outcome would have been different had plaintiffs originally requested the information instead of waiting.

D) Spreadsheets

In evaluating the necessity of metadata to the spreadsheets, the court in *Aguilar* explained that the relevance of the metadata would depend on its "complexity and purpose."[74] Where the metadata contains formulas necessary for an understanding of the information in the spreadsheet, for example, this information would be relevant. In *Aguilar*, the court did not find the metadata particularly necessary to an understanding of the information contained in the spreadsheet. The spreadsheets "merely list[ed] the date of a particular operation, the field office that conducted the operation, the number of arrests made, and a breakdown of those arrests into different categories."[75] Nevertheless, the court found that the request for metadata was "not unduly burdensome" and thus ordered production of metadata for the spreadsheets.

Similarly, in *Williams v. Sprint*, an employee, on behalf of 1,727 other plaintiffs, alleged that age was a determining factor in her employer's decision to terminate her employment during a reduction in workforce. Plaintiff requested production of the spreadsheets used to arrive at the decisions to terminate. In response to the court's order to produce, the employer stated that it provided the spreadsheets requested. However, the formulas behind the spreadsheet calculations had been scrubbed or deleted from the spreadsheets produced to plaintiffs. Defendants argued that the formulas that came with the spreadsheets were not relevant. The court did not agree. It ruled that the employer failed to show cause why it should not produce the electronic spreadsheets in the manner in which they were maintained. The defendants were ordered to produce them as they were maintained. The court ruled that the employer avoided sanctions, this time, by its decision to voluntarily reproduce "unlocked" versions of the spreadsheets.[76]

E) Audit Trails

In *Aguilar*, the plaintiffs also "[sought] metadata associated with certain documents that the Defendants…produced from hierarchical databases."[77] The court noted that "audit trail information [was] available and would enable the Plaintiff to determine what changes were made and when."[78] Here it was not clear whether any information found in the

---

[74] *Id.* at 354.
[75] *Id*. at 362.
[76] 240 FRD 640.
[77] *Aguilar v. Immigration & Customs Enforcement Div.,* 255 F.R.D. 350. See Appendix for explanation of Hierarchical Databases.
[78] See Appendix for explanation of Audit Trails.

metadata would have been useful to plaintiffs or relevant to the case. In order to show they had not perpetuated fraud, the defendants agreed, at the suggestion of the Court, to show the plaintiffs how the databases functioned. If they failed to do this within a reasonable time, the court explained, the defendants would have to make the metadata available to the plaintiffs.

F) File Sharing

Another recent case expanded the scope of discoverable metadata, allowing parties more resources with which to prove their cases. In *Maverick Recording Co. v. Harper*[79] the plaintiff recording company sought metadata to prove the defendant was responsible for file sharing. The plaintiff was able to do this by tracing the defendant's Internet protocol address. "The company captured screen shots showing all of the files that [defendant] was sharing. It also captured the metadata associated with each file, which included the name of the artist and song."[80] The plaintiff also "initiated downloads of the audio files to verify their existence and recovered metadata from which it could identify the artist and song title of each file."[81] Just as the plaintiffs in *Aguilar* needed metadata to show who sent and received emails, the metadata here showed who downloaded the music files and to which computer.

G) Electronic Photographs & Images

The court in *United States v. Welton* found critical uses for metadata to authenticate various electronic photographs and images.[82] In this case a stack of photos depicting child pornography was found hidden in the nursery of a church. The defendant did not deny that the pictures found belonged to him. He even admitted that he had added to the stack using online links. For purposes of establishing the extent of his crimes, however, the court looked to electronic evidence in the case. In examining the evidence, the court considered the information found in the exhibits. Because different dates were found based on the different formats of the exhibits, the court found different browsers were used to download the images. The court also found that because the documents bear different time and date metadata, they were printed out at various times.[83] The court distinguished between using the metadata to establish what dates and times the documents were printed out and, as they actually did in this case, using the metadata to prove the documents were printed out at different times. In *Welton*, defendant's timing in accessing the documents supported "a finding of knowing possession," an element of the crime charged. Next the court addressed interstate commerce and considered whether the images were transmitted over the Internet. The court relied on the government expert's testimony "that metadata in the printouts constituted circumstantial evidence that the images were printed from the Internet websites."[84]

---

[79] *Maverick Recording Co. v. Harper*, 598 F.3d 193 (5th Cir. Tex. 2010).
[80] *Id*. at 194.
[81] *Id. at* 196.
[82] *United States v. Welton*, 2009 US Dist. Lexis 110657 (C.D. Cal. 2009).
[83] *Id*. at 18.
[84] *Id.* at 73.

Based on the expert testimony of the government witness, which was based on the metadata of the photographs, the court found that the Government "proved beyond a reasonable doubt that defendant knowingly possessed child pornography" in violation of § 2252A(a)(5).[85] The court also held, however, that the Government did not prove beyond a reasonable doubt that defendant knowingly *received* child pornography in violation of § 2252A(a)(2).[86] There was enough evidence to establish possession of the child pornography by the defendant, but the metadata was not enough, in this case, to show that the pornography was received or distributed by the defendant.

Welton shows the limits to metadata as a means of authenticating records. There the court found the metadata sufficient to show the defendant viewed the content in question, but insufficient to show that he received or distributed the content. While metadata can be an effective tool for authenticating records and documents, it is not without limits.

H) Public Records

In *Lake v. City of Phoenix* the Arizona Supreme court ruled that metadata associated with public records are indeed a part of the public record itself.[87] The plaintiff, who was a police officer, had lodged an administrative complaint and eventually a federal lawsuit regarding his job performance. The employee had also submitted a public records request to the city, seeking notes kept by his supervisor documenting the employee's work performance. After reviewing copies of the notes, the employee suspected that they had been backdated when prepared on a computer. The employee requested metadata contained inside the supervisor's notes file. The city denied the employee's public records request for metadata in the electronic version, claiming the metadata was not a public record. The supreme court disagreed with the city's contention. It held that if the city maintained a public record in an electronic format, then the electronic version, including any embedded metadata**,** was subject to disclosure under Arizona's public records laws, pursuant to Arizona Revised Statute, Section 39-121 of 2001. The Arizona Supreme Court concluded that when a public officer uses a computer to make a public record, the metadata becomes part of the document as much as the words on the page.

Lake joins at least one other case in Washington where it has been held that metadata is indeed part of the public record itself. [88] Therefore, government entities must ensure the capture of metadata not just in the event of litigation, but also for public records requests.

---

[85] 18 USCS § 2252A(a)(5) (2010).
[86] 18 USCS § 2252A(a)(2) (2010).
[87] 218 P.3d 1004 (AZ Sup. 2009).
[88] *See O'Neill v. City of Shoreline*, 187 P. 3d 822 (Wash App. 2008) (where the Washington Court of Appeals held that metadata contained in emails received by the Mayor in her personal account and referenced at a city council meeting were part of the public record).

IV. DISCERNIBLE PATTERNS FOR RECORD RETENTION & DISCOVERY PRESERVATION

This paper set out to identify metadata fields for best information governance practices, including records and information management (RIM), as well as e-discovery preservation obligations. Records and information governance professionals should understand that records captured for records management purposes differ from data preserved in anticipation or for pending litigation. The former is limited to records used to document the business transactions of the organization. The latter is, of course, anything existing in the systems, irrespective of the records retention schedule,[89] that is relevant to the subject matter of the anticipated, threatened or pending litigation or agency investigation.[90]

A) Metadata Fields to Preserve in Declaring "Records"

The court decisions and think tanks discussed above address the growing list of issues e-discovery presents and the way courts have thus far dealt with these issues. In the absence of clear statutory law outlining the proper procedure for how to manage and store such evidence, companies must be proactive and vigilant with their records. The first step is the implementation of a clear and exhaustive records retention policy. A clear policy is one that provides guidelines for creation, storage, and preservation. Preservation of metadata should include, at minimum, system and embedded metadata. Having such a policy in place will help ensure compliance with the rules and also will reduce the resources, like time and money, spent in anticipation of litigation.

Usability of the record is a key factor for determining what metadata to preserve. The court in *Aguilar* addressed the applicability of FRCP Rule 34's reference to producing the information in "reasonably usable form."[91] It noted that other courts had recognized this to mean that searchability of the document should *not* be limited.

Accordingly, a records retention policy and procedures should focus on preserving types of metadata as follows:

Substantive metadata (i.e., reflective of substantive changes to a document) should be associated with drafts of a record, which most companies define as a non-record. It does not serve much use once the ultimate record is declared. Thus, substantive metadata could be considered a non-record. It should not be necessary to preserve it in declaring the ultimate record. Similarly, metadata fields that are frequently updated, such as last opened dates, are not considered discoverable in most cases, as noted in the Seventh Circuit's Pilot Program. It is thus safe to say that it does not need to be captured as part of the record.

On the other hand, system metadata (e.g., file names and extensions, sizes, creation dates) should be preserved. This data is crucial to authenticating the record whether in

---

[89] See attached Appendix for definition of a Retention Schedule.

[90] *See*, Isaza, John & Jablonski, John, *7 Steps for Legal Holds of ESI and Other Documents*, Introduction, pages 2-4 (ARMA International 2009).

[91] *Id.* at 358.

litigation or for any other purpose where authenticity of the record is paramount, like archival or historical uses.

Finally, embedded metadata (i.e., content data contained in a native file like formulas or hidden columns) is arguably relevant from both a records management point of view and in discovery. The courts uniformly agree that it is crucial to understanding a document. Therefore, it should be preserved in declaring the record.

B) Metadata Fields to Preserve in Discovery or Threatened Litigation

Although there are some advocates who propose that metadata should be presumed not accessible in discovery, the fact is that the current think tanks and case law essentially dispute this position. Several fields of all three types of metadata (substantive, system and embedded) should be preserved for e-discovery purposes (i.e., when litigation is either threatened or pending). Substantive metadata, for one, could be relevant. Substantive changes to a document, such as prior edits or editorial comments, may tell a story about the document that goes beyond the document and into the case at hand. It may not be necessary to declare and classify a record, but it should be preserved as relevant to litigation or an agency investigation that is pending or threatened.

As far as system metadata (e.g., file names and extensions, sizes, creation dates), this data could be crucial to authenticating the record in litigation. This is the kind of information that goes to the heart of whether a document is what it was purported to be, as discussed in the *Lorraine* case. Thus, for discovery purposes, system metadata is highly relevant. As noted in *Aguilar*, it aids with access, search and sorting of documents.

Similarly, embedded metadata is critical to interpreting the document. Formulas, hidden columns, etc. are all part of the document in its entirety. Its preservation should not be discretionary. Besides that, these fields are not easily changed, so there is no excuse for failure to preserve them for threatened or pending litigation.

On the whole, for discovery preservation purposes, the scope of metadata to preserve may need to be done on a case by case basis, absent a policy of retention and production of all data in native format – a potentially risky or unrealistic path. Courts are looking for parties to preserve information that would render the document "usable." What is reasonably usable depends on the particular circumstances of a case. According to the Sedona Principles, some factors counsel should consider include: "(a) the forms most likely to provide the information needed to establish the relevant facts of the case; (b) the need for metadata to organize and search the information produced; (c) whether the information sought is reasonably accessible in the forms requested; and (d) the requesting party's own ability to effectively manage and use the information in the forms requested." Due consideration should be given to the form in which records are ordinarily maintained.

C) Minimum Fields to Preserve Based on or Implied in Case Law

Case law indicates that the following "reasonably accessible metadata" should be considered, at minimum, when determining what metadata fields to preserve. These minimum guidelines should be applied for both records management purposes and also for discovery purposes.

1. **For Emails**

   Metadata fields that tend to show who knows what, when such as:

   - Dates created and modified
   - Subject
   - Size of file and attachment
   - Modified by
   - Senders and receivers
   - Attachments
   - Author
   - Custodian
   - Beginning document page number
   - End document page number
   - Beginning attachment page number
   - End attachment page number
   - Hyperlinks

2. **For documents such as WORD, WordPerfect, and Spreadsheets**

   Consider the complexity and purpose of the document. Although this request sounds simple, for records managers this poses a nightmare. Therefore, a uniform approach to all of these documents should be considered. Again, preserve all metadata fields that render all documents created in these applications usable. Fields to consider might include:

   - Dates created and modified
   - Authors and modifiers (the latter not necessary to declare records)
   - Custodian
   - Size of file
   - Beginning document page number
   - End document page number
   - Key formulas embedded in any of the documents
   - Hyperlinks

   Specifically regarding spreadsheets, where the document contains formulas necessary for an understanding of the information in the spreadsheet, this information would be relevant. As noted in *Aguilar*, in the case of a spreadsheet the need for metadata "depends upon the complexity and purpose of the spreadsheet."

3. **For Electronic Music File Shares, Photos, Animation, Videos, Graphics & Images**

   Similarly to emails noted above, electronic music, photos, animations, videos, graphics and images require at minimum the retention of who, what and when types of information. These should include:

- Folders
- File Names
- Author (if known)
- Dates created or downloaded
- Date modified (not needed to declare records)
- Modified by (if known, and not needed to declare records)
- Custodian
- Formulas embedded in any of the documents, if any
- Size of file
- Hyperlinks, if any

CONCLUSION

Records and information governance professionals should understand that records captured for records management purposes differ from data preserved in anticipation of or for pending litigation. The former is limited to records used to document the business transactions of the organization. The latter is, of course, anything existing in the systems, irrespective of the records retention schedule, that is relevant to the subject matter of the anticipated, threatened or pending litigation or agency investigation.

The current preference advocated by some legal experts and courts is to produce information in native format, which by extension includes the relevant metadata without separate preservation steps. However, this preferred method has its pitfalls, especially for organizations where records are migrated into another format at some point or another. For such organizations, it is imperative to have procedures or, at minimum, guidelines that establish what metadata to retain in declaring and classifying the organization's official records. This will be the organization's first line of defense when questioned regarding record retention practices and ultimately regarding preservation for discovery purposes. Of course, once litigation is threatened or pending, organizations should strive to preserve all metadata that is reasonably accessible and that would render the relevant documents usable. Once discussions commence with opposing counsel, further details on precisely what metadata to produce may come to light.

In discovery, courts are putting a great deal of pressure on the requesting party to identify early on what metadata fields to preserve and produce. Thus, if organizations preserve certain fields as a matter of course and in their business practices, once litigation commences the onus will be on the opposing side to specify any additional fields to preserve for production purposes. If the fields exist, then they must be produced, if reasonably accessible. If they do not, the retention policy will explain to the court why the fields are no longer available. Ultimately, in determining what to keep for the "record" or in anticipation of pending or threatened litigation/investigations, the ultimate test will be whether the metadata preserved aids in rendering the document usable.

APPENDIX: GLOSSARY & EXPLANATION OF KEY IT, LEGAL AND RIM TERMS USED

Application - Application software, also known as software application, application or app, is computer software designed to help the user to perform a singular or multiple related specific tasks. Typical examples are word processors, spreadsheets, media players and database applications.
http://www.en.wikipedia.org/wiki/Software_Application, downloaded from *Wikipedia* on September 2, 2010.

Audit Trails - The Dictionary of Storage Networking Technology defines an audit trail as "A chronological record of system activities that enables the reconstruction and examination of a sequence of events and/or changes in an object. The term audit trail may apply to information in an information system, to message routing in a communications system, or to any transfer of sensitive material and/or information." http://www.snia.org/education/dictionary/a/, downloaded August 30, 2010.

Authentication – In the law of evidence, the act or mode of giving authority or legal authenticity to a statute, record, or other written instrument, or a certified copy thereof, so as to render it legally admissible in evidence. *Black's Law Dictionary*, Abridged 5th Edition (West 1983).

Discovery – The required disclosure of relevant items in the possession of one party to the opposing party during the course of legal action. *ARMA Glossary of Records Management and Information Terms*, 3rd Edition (ARMA International 2007).

ERP System – Enterprise Resource Planning Systems (ERP) are defined in *PC Magazine.com* as follows: An integrated information system that serves all departments within an enterprise. Evolving out of the manufacturing industry, ERP implies the use of packaged software rather than proprietary software written by or for one customer. ERP modules may be able to interface with an organization's own software with varying degrees of effort, and, depending on the software, ERP modules may be alterable via the vendor's proprietary tools as well as proprietary or standard programming languages.

An ERP system can include software for manufacturing, order entry, accounts receivable and payable, general ledger, purchasing, warehousing, transportation and human resources. The major ERP vendors are SAP, Oracle (PeopleSoft and J.D. Edwards), SSA Global (Baan) and Microsoft. Lawson Software specializes in back-end processing that integrates with another vendor's manufacturing system.
http://www.pcmag.com/encyclopedia_term/0,2542,t=ERP&i=42727,00.asp, downloaded October 31, 2010.

Hierarchical Databases - Hierarchical databases are defined in *PC Magazine.com* as follows: A database organization method that is structured in a hierarchy. All access to data starts at the top of the hierarchy and moves downward; for example, from

customer to orders, vendor to purchases, etc. Contrast with relational database and network database. http://www.pcmag.com/encyclopedia_term/0,2542,t=hierarchical+database&i=44239,00.asp, downloaded August 20, 2010.

Metadata - Stated simply, metadata is data about data. ARMA International defines metadata as "Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource." *ARMA Glossary of Records Management and Information Terms*, 3rd Edition (ARMA International 2007).

Migration - As to migration to another format, the *Tech Terms Dictionary* goes on to explain: "When you use the "Save As..." command to save a file, you may be given the option to save the file in a different format. For example, you might be able to save a Word document as a plain text (.txt) file or a rich text (.rtf) file. These formats are not native to Microsoft Word, but can still be opened by the Microsoft Word program. Similarly, Adobe Photoshop saves files as Photoshop documents (.psd files), but can also save them in .jpg and .gif formats, among others. It is usually best to save a file in a program's native file format because you can be sure it will store all the data you have created with the program. While other formats may be more compatible with other programs, they might not save all the information in the file. For example, if you save a Word document as a plain text file, all the text formatting you had added will be removed. Saving a Photoshop document in JPEG format will reduce the image quality and flatten all the image's layers." downloaded from http://www.techterms.com/definition/nativefile on August 20, 2010,

Native Files - Native files are explained in the *Tech Terms Computer Dictionary*, downloaded from http://www.techterms.com/definition/nativefile on August 20, 2010, as follows: "When you save a file using a certain program, the file is often saved in a proprietary format only that program can recognize. For example, if you save a Microsoft Word document, it is saved as a Word document (i.e. mydocument.doc). This is a native Word file -- that is, the file format is native to the Microsoft Word application and may not be recognized by other programs."

Retention Schedule – A comprehensive list of record series, indicating for each the length of time it is to be maintained and its disposition. *ARMA Glossary of Records Management and Information Terms*, 3rd Edition (ARMA International 2007).

Unstructured Data – Unstructured data is defined in *SearchBusinessAnalytics.com*, downloaded from http://searchbusinessanalytics.techtarget.com/definition/unstructured-data on September 8, 2010, as follows: "Unstructured data is a generic label for describing any corporate information that is not in a database. Unstructured data can be textual or non-textual. Textual unstructured data is generated in media like email messages, PowerPoint presentations, Word documents, collaboration software and instant messages. Non-textual unstructured data is generated in media like JPEG images, MP3 audio files and Flash video files."

Funds for this study were provided by the

The ARMA International Educational Foundation is the non-profit, (501(c)3, affiliate of ARMA International, the primary professional association for the records and information profession in the world.

Mission
The ARMA International Educational Foundation supports education and research initiatives that promote the advancement of both information managers and the information management profession. Recorded information is the lifeblood of the modern organization, but rarely is it treated as a critical asset, primarily because there is little quality research to create the comprehensive body of knowledge required to support information management as a profession. The AIEF purpose is to answer that need by soliciting funds for this research and then providing a vehicle through which conclusions can be tested, documented and communicated to the information management community.

If you found value in this publication, please consider making a financial contribution to the Endowment Fund of the Foundation. This can be accomplished by visiting the Foundation's web site, www.armaedfoundation.org, or by contacting

Foundation Administrator
ARMA Int'l Educational Foundation
1609 Terrie Drive
Pittsburgh PA 15241
USA

Additional information about the Foundation can be found at GuideStar
The National Database of Non-profit Organizations
http://www.guidestar.org/search/report/gs_report.jsp?ein=31-1556655

Comments about this publication and suggestions for further research are welcome. Please direct your inquiry to the Foundation Administrator.