

EXAMINING METADATA: Its Role in E-Discovery and the Future of Records Managers

Recent developments in law, standards, and technology suggest that metadata will play an important part in the evolution of electronic records management.

Julie Gable, CRM, CDIA, FAI

The need to produce metadata for e-discovery, the emergence of international best practices regarding metadata management, and the development of metadata repositories all suggest that records managers need to prepare themselves for a new role.

Metadata and E-Discovery

Aguilar v. Immigration and Customs Enforcement was a class action suit alleging unlawful searches and seizures of plaintiffs' homes. At first, Aguilar did not ask for metadata, but after Immigration and Customs Enforcement (ICE) had completed its e-discovery, Aguilar asked for e-mail and other electronically stored information with corresponding load files, which contain metadata fields and extracted text.

ICE replied that the request was burdensome and said it would produce metadata only for documents where the plaintiffs could demonstrate its relevance to their claims. After much research, the court ordered ICE to produce metadata for e-mail, Word, PowerPoint, and Excel files, but it shifted some of the cost to the plaintiff.

Aguilar v. Immigration and Customs Enforcement is interesting to lawyers because the court compelled the production of metadata after initial discovery – in effect, allowing Aguilar to “double-dip” with its e-discovery request. But more interesting are the steps the court took to reach its decision because they show an increasing knowledge and sophistication regarding metadata.

First, the court defined three kinds of metadata (see sidebar on page 30).

1. *Substantive* metadata is application-based and may contain modifications, edits, comments, etc., that were not necessarily intended for adversaries to see. Much has been written about

the ethics of examining substantive metadata if it has been supplied inadvertently.

2. *System-based* metadata includes information automatically captured by the computer system, such as author, date, time of creation, and date of modification.
3. *Embedded* metadata consists of text, numbers, and content that is directly input but not necessarily visible on output, such as spreadsheet formulas or hyperlinks.

In *Aguilar*, the court consulted three sources: *The Federal Rules of Civil Procedure* (FRCP), the Sedona Principles, and case law. Although metadata is not directly addressed in the FRCP, it is subject to the rules of general discovery if it is relevant to the legal matter and is not considered privileged. The FRCP's balancing test also applies to metadata, namely, it asks if the probative value of the metadata is worth the cost to produce it, and if so, who should bear the cost.

The Sedona Conference's® *Sedona Principles for Electronic Document Production* offer two primary considerations regarding metadata in evidence production:

- 1) Is metadata relevant?
- 2) Does it enhance the utility of the documents, including the ability to search them?

Note that principle 12, which concerns metadata, has changed from the 2004 version published in *The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production*, which stated, “Unless it is material to resolving the dispute, there is no obligation to preserve and produce metadata absent agreement of the parties or order of the court.”

In the 2007 version published in *The Sedona Principles: Addressing Electronic Document Production, 2d ed.*, principle 12 states that the form of production should take into ac-

count “the need to produce reasonably accessible metadata that will enable the receiving party to have the same ability to access, search, and display the information as the producing party where appropriate or necessary in light of the nature of the information and the needs of the case.”

Finally, case law provides precedents for compelling metadata production. In *Williams v. Sprint* (D. Kan. 2005), the court ruled that presumption is in favor of producing metadata unless the producing party objects, the parties agree otherwise, or the court issues a protective order.

In *Nova Measuring Instruments Ltd. v. Nanometrics, Inc.* (N.D. Cal. 2006), the magistrate granted a motion to compel the defendant to produce electronic documents in their native format with all original metadata where the defendant originally agreed to do so and provided no reason why it could not. In *re NYSE Securities Specialists Litigation*, (S.D.N.Y. June 14, 2006), the court ordered all electronic documents to be produced in native format.

Metadata's Usefulness

Lawyers want metadata because it can be used to show things not evident from the document's content alone. For example, a full e-mail header indicates when a message was sent and when it was received, as well as the Internet protocol addresses of the computers used to send and to read the message. Comparing a document's system metadata and substantive metadata may be used to show that a document was tampered with.

In disputes involving commissions or bonus amounts, it could be useful to show embedded metadata regarding formulas used to do calculations. The important point is that metadata is valuable because it can be used to show contextual, processing, and use information associated with a document. It can also illustrate the document's chain of custody, storage

TYPES OF METADATA

Substantive Metadata

- Created by the application
- Part of the file itself and moves with the file
 - Example: MS Word's "track changes" feature allows comments, edits, changes, etc. to appear in documents copied or sent to others. Default settings for Office 2007 programs automatically display these when documents are opened or saved.

System-Based Metadata

- Created by the computer's operating system to track the file
- Stored external to the file itself
- Stored demographics about the file, e.g., name, size, creation, modification, usage
 - Every active file has at least one corresponding block of system metadata
 - MS Word, PowerPoint, and Excel have 80 easily accessible substantive and system metadata fields tracked
- Can be unintelligible outside its native environment, so it must be labeled
- May not be text – for example, a flag in a database

Embedded Metadata

- Stored within the file, but not necessarily visible when the file is displayed or printed. A kind of third dimension that is layered into the document beneath the onscreen data that is not visible.
 - In spreadsheets, examples are formulas, linked values, pivot tables.

locations, access history, and versions.

Experts caution that metadata is useful but fragile. Opening or previewing documents will alter their last access date. In some systems, copying a file changes the creation date on the copy to the date it was copied. Saving a file can alter its last modified date. Some applications carry the original author's name even if another person copies the file and modifies it, and some antivirus applications actually "touch" every file. If there is a chance that metadata will be needed for litigation, it is imperative not to "peek" just to see what exists because doing so will alter the metadata.

The E-Discovery Reference Model from Socha-Gelbmann (see Figure 1 on page 31) shows the processes associated with electronic discovery. In the model, information management appears as a means to reduce the volume of information that needs to be produced, but the question is, will RIM – should RIM – expand to cover preservation and collection, particularly with regard to metadata? Is metadata a record that needs to be managed?

More and more, the answer is a thundering "yes" from the worlds of law, standards, and technology. *The Sedona Best Practices Guidelines & Commentary for Managing Records & Information in an Electronic Age* (SBP) notes that metadata allows entities to better retain and organize information. Metadata, such as title, author, date created, date finalized, and the software application version used to create a record, are all considered useful for organization.

SBP cautions that metadata retained in the due course of business "may be discoverable in its complete and original form." SPB discusses preservation and migration of metadata over time, noting that technical, descriptive, and preservation metadata may be needed to show how a record was created, maintained, and related to other records. To survive

migrations, SBP notes that records will need metadata that enables them to exist independently of the system used to store and retrieve them, a fact not lost on electronic records management technology companies.

Metadata Standards

Metadata's usefulness extends well beyond legal purposes, however. The international standard *ISO 23081-1: Metadata for Records* is a guide for understanding, implementing, and using metadata within the framework of *ISO 15489-1 Information and Documentation – Records Management – Part 1*.

ISO 23081 addresses the relevance of records management metadata in business processes, defines the roles and types of metadata needed for records management processes, and sets a framework for capturing metadata.

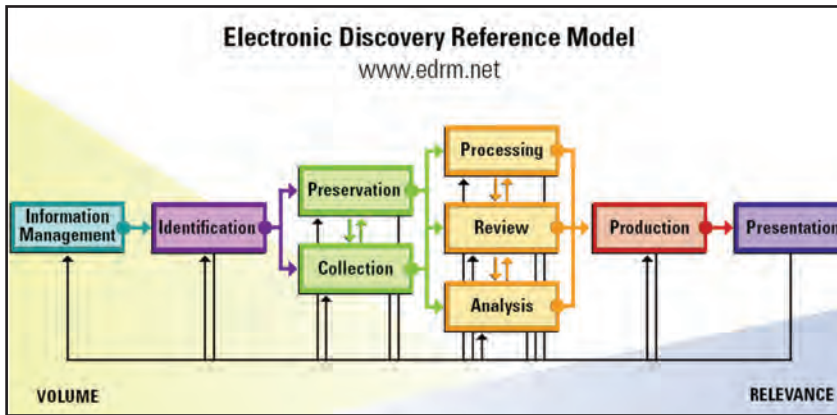
It states that the context for a record's creation is the business process it is related to. Metadata should capture the agents involved in the record's creation, as well as the record's content, appearance, structure, and technical attributes.

Record structure can refer to the record's physical structure, its technical structure, or its logical structure. Physical structure might describe media, such as paper; technical structure might describe a file format such as PDF; logical structure, however, recognizes the relationships between the data elements that comprise the record.

Notice that by mentioning logical structure, *ISO 23081* acknowledges that records are not just documents and e-mails as we know them. Records may also be the result of electronic pieces brought together from several applications and assembled for presentation, viewing, etc. Such pieces could include video, animation, and sound recordings, among other things.

Of interest to records managers is that *ISO 23081* defines a metadata

Figure 1: : Electronic Discovery Reference Model



Source: www.edrm.net

record and discusses who is responsible for it. First, the standard states: “Metadata about the record and metadata accruing in its management form a metadata record which must be managed.”

Metadata records fall into several categories based on the purpose of the metadata. Examples from the standard include:

- Metadata accruing from e-business transactions
- Metadata regarding record preservation
- Descriptive information regarding an information resource
- Retrieval metadata that is used to facilitate search
- Rights management metadata that describes permission to use resources

Next, the standard assigns retention for metadata, noting, “It is essential to keep this metadata record at least as long as the original record exists.”

Lastly, and perhaps most importantly, RIM professionals are given responsibility for the reliability, authenticity, usability, and integrity of metadata associated with records. What is more, records managers’ role is to participate in defining metadata requirements, develop metadata policies and strategies, and monitor metadata creation. *ISO 23081*, like

all standards, is not mandatory, but by offering guidance, it also provides insight as to what the future may bring.

The emphasis on metadata can also be seen in the latest version of the *Model Requirements Specification for the Management of Electronic Records (MoReq2)* which defines generic requirements for an electronic records management system (ERMS).

MoReq2 provides a metadata model that has two related purposes. First, metadata as specified in the model becomes a method to allow the exchange of records between ERMSs without any loss of functionality. Second, MoReq2’s metadata model provides a basis for developing an XML schema, that is, a standard way to describe metadata elements in terms of their structure and content.

MoReq2’s metadata model describes a minimum set of metadata elements that an ERMS must be able to export, import, and process. An “element” in MoReq2 parlance is the field used to hold a metadata value.

Examples of metadata elements for an ERMS would include classification schemes, record types, components, retention and disposition schedules, users, groups, and roles. The MoReq2 metadata model is meant to be consistent with *ISO 15489*, *ISO 23081*, and *ISO 15836, Information and Documentation – The Dublin Core*

Metadata Element Set for describing information resources.

Metadata and Technology

Exactly how records management technology will cope with the new demands for metadata remains to be seen, but some ERM software firms are recognizing that metadata repositories may hold the key to managing electronic records in place, that is, within the records’ native application.

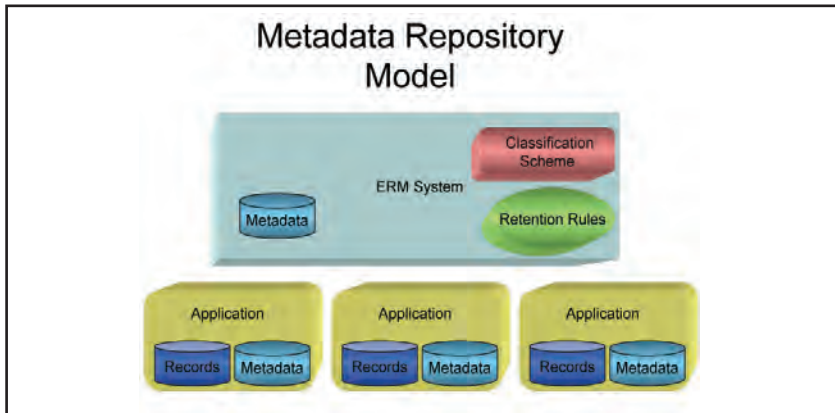
This approach does not seek to build an electronic archive containing the records themselves; rather the records remain in the applications that create or manage them, but relevant metadata about the records – such as that needed for description, retrieval, policy enforcement, interoperability, and preservation – is managed and maintained in a metadata repository that is separate from where the records themselves actually reside. (See Figure 2 on page 32.) The benefits of this architecture are that it can scale, easily handling the massive volumes of information typically found in large organizations – which can range from hundreds to thousands of terabytes scattered across mainframes, servers, and network drives. The metadata repository also offers a single point of management for the metadata record itself, thereby facilitating policy management, retrieval, and metadata for e-discovery.

Because the metadata repository could also contain whatever data is needed for interoperability, it has the potential to allow records to exist independently of the system used to store and retrieve them. Several companies have already begun to offer a metadata repository approach to records management and information.

The Future

Clearly, metadata has become far more than just “data about data.” Metadata can be used to describe information assets, enhance retrieval, enforce policy, show ownership,

Figure 2: Metadata Repository Model



demonstrate authenticity, foster interoperability, and enhance electronic discovery.

But there's more: beyond records management, IT looks at metadata as a means to foster integration of diverse applications, a way to cull and relate information from data silos, a challenge currently faced by electronic patient health records. For systems integrators, metadata represents a way to knit disparate views of a company's data together for business intelligence and productivity, as well as for information governance purposes.

Managing metadata is the direc-

tion of the near future, particularly as content management, records management, and e-discovery systems converge and consolidate. Deciding what metadata to keep will depend on the needs of a diverse set of interested parties in legal, compliance, records management, information technology, and business functions. Records managers will find themselves becoming metadata managers, the culmination of a prediction made years ago by Iron Mountain's Executive Chairman of the Board Richard Reese that, "In the future, records managers will be database managers."

Metadata management will focus more on the context, relatedness, and use of records than records management does today. Farther in the future, as documents themselves morph into containers of video, graphic, sound, and text components, metadata will be the glue that holds these diverse pieces together.

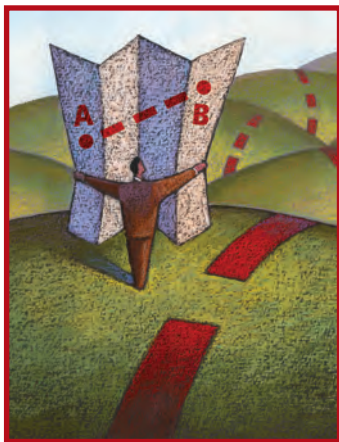
Traditional, document-centric records practices likely won't translate well in the world of e-component management. But tomorrow's metadata managers may find themselves doing far more satisfying things than managing for crises in e-discovery and compliance. They'll have a pivotal, recognized role in helping organizations to relate information for greater intelligence, productivity, and performance.

Given the growing emphasis on metadata, the future is closer than it appears. **ENB**

Julie Gable, CRM, CDIA, FAI, can be contacted at julie.gable@verizon.net. See her bio on page 62.

See references for this article at www.arma.org/imm/sept09/metadata.pdf.

Where's Your Career Headed?



Now there's a tool to guide you, available only from ARMA International:

The RIM Self Assessment

Regardless of what stage you're at in your career, this online tool will help you identify your strengths and any gaps in your skills and knowledge. It will even point you to resources that can help you address those gaps.

The tool is easy and it's **free** to members of ARMA International. And your results will remain confidential. Start your RIM Self Assessment today!

www.arma.org/competencies 