XML Marks the Future for Electronic Records

Learning about extensible markup language allows RIM professionals an opportunity to determine how electronic records will be stored and searched

Roger Winters

Thanks to extensible markup language (XML) – and its relatives and derivatives – there are powerful new possibilities for electronic records. Records and information management (RIM) professionals need to learn about XML and what it can do for their organizations. By investing the time in such education, RIM professionals can ensure that records based in XML technology – increasingly the case – are properly managed.

While it is becoming an accepted standard, most RIM managers will never "do" XML although it is something they need to understand if not master technically. In fact, looking at the technical side of XML may quickly turn many RIM professionals off since one risks being overwhelmed by acronyms and jargon. By keeping a clear focus on business issues, however, RIM professionals need not be intimidated by the language and concepts of XML.

Our Electronic (Paper) Records

RIM managers increasingly adopt electronic documents rather than hard-copy records to gain efficiencies. They have embraced digital images, creating electronic graphic copies of paper documents because they are easier to transmit, store, and access than the paper records they represent. This choice accounts for the present challenge to manage the exponential growth of electronic records. In many ways, however, RIM managers have tried to treat e-records like paper records (e.g., printing out e-mail messages and filing them in file cabinets). Most electronic records do, in fact, resemble paper records, whether word-processed documents, database reports, or e-mail messages. Printing and using the Adobe portable document format (PDF) are seen as ways to "lock" electronic documents so they appear safe from unauthorized modification. These items can be managed as records because it is known they could be printed onto paper if need be and handled as hard copies. PDF is widely used because it retains the appearance of the "original" paper document. Many courts and businesses accept or require that electronic filings be provided in PDF.

Humans need to see written information displayed as words on a surface, whether on paper or projected on a screen. Paper usually is the surface of choice for reading. Despite prophecies of a "paperless office," paper is used more than ever before, particularly as a way to present information.

At the Core

This article

- Provides an explanation of what XML is and does
- Describes how XML may change electronic records
- Urges RIM professionals to help build standards for XML use

What is XML?

"XML" has a general meaning; its particular uses must be explained. There can be forms of XML for any type of business or activity. There is XML for chemistry, for pharmaceuticals, for webpage building, and there is – or soon will be – XML for contracts, court filings, and insurance. Sellers of many products proudly

claim they "have XML" or "use XML" in their "solutions." They do not always explain what that means.

Often a vendor, technologist, or implementer talks about how XML figures in a plan, product, or service, leaving out information really useful to RIM professionals. Saying that a solution uses XML is similar to saying that something uses electricity. In the same way that electricity in a lightning bolt is not the same as that in a cyclotron, or in an MP3 player, a car battery, or a street lamp, what it means when someone says a solution "uses XML" can vary greatly.

Simply put, XML is a markup language. In 1974, Charles F. Goldfarb, a lawyer, invented the standard generalized markup language (SGML) later adopted as an international standard (ISO 8879). XML and its better known relative, hypertext markup language (HTML), are both derived from SGML. (XML 1.0 became a World Wide Web Consortium [W3C] recommendation on February 10, 1998.)

A markup language is not software. It is a meta-language, which is defined as a language used to describe or analyze a language. More widely known than XML, HTML is a meta-language with the specialized purpose of describing how webpages are to be structured when a browser displays them. An XML meta-language describes how the meanings of items of data are to be interpreted so they can be processed by applications.

With "markup," clearly defined codes are placed in a document; text is typically intermixed with the markup codes so they are all in the same data stream or file. Most RIM professionals have already used markup to edit hard copy (e.g., using the symbol ^ to indicate inserted text). A person or a program then uses the markup codes as instructions for how to process the marked items.

In the electronic document, markup can variously affect the handling of an item's:

1. **Format** – directing how the information or element is to be displayed
2. **Structure** – signaling how to treat the item as a title, chapter, sub-heading, table, illustration, footnote, appendix, etc.
3. **Meaning** – indicating how to interpret the item based on the tag name's stipulated definition, the item's attributes, and its relation to other data/information

In a sense, one doesn't use XML itself. Instead, XML is the rulebook from which particular meta-languages are created.

The first two types of markup are generally handled in HTML documents. XML is sometimes used for structure, but it is mainly used to indicate the semantic meaning of the tagged data elements. A markup tag works like a "toggle," or a switch – it turns "ON," then "OFF." Markup tags are usually paired, with the first indicated by the tag name placed inside **angle brackets** – <FirstName> – with the paired closing tag the same except for a forward slash preceding the tag name: </FirstName>. Within a given text file, we might find (at various places) <FirstName>Betsy</FirstName>… <FirstName>Jim</FirstName>… <FirstName>George</FirstName>… and so forth. A program written to create a list of names found in a document could parse it, locate appropriate tags, and re-use the data items, for example, by displaying the following list:

First Names from the Document
Betsy
Jim
George

Markup tags do not themselves format, impose structure, or interpret data: a browser or a parser interprets tags and acts on them accordingly. Once tagged, a software application can interpret and use the data elements elsewhere, in other documents, computers, applications, and so forth. This is **re-purposing** data. Re-purposed data can be used by incompatible systems without having to move data by hand (the "read and re-type" method) by using the same tags for the same things. Having clear standards in tagging data elements makes information exchanges among different systems possible. Standard tags, in a sense, turn a document into a container holding reusable data elements. The document may be "frozen" so it will not itself be changed, but the data within it could be copied and reused endlessly.

XML Rules and Specifications

In a sense, one doesn't use XML itself. Instead, XML is the rulebook from which particular meta-languages are created. Strictly speaking, to identify or assign meaning to content, a semantic markup language based on XML is used. (The strict rules for XML are found in the XML 1.0 specification adopted by the W3C. To see the XML 1.0 specification, go to *www.w3.org/TR/REC-xml/*. A somewhat more user-friendly annotated XML specification is at *www.xml.com/axml/testaxml.htm.*)

XML is strict and unforgiving. That can be a strength and also a problem. Any of the XML tags here – <LastName> // <lastName> // <LASTNAME> // <last-name> – could be designated as meaning "a person's surname," but in XML, each is totally different because XML is case sensitive. Diverse systems, software, and product names, etc., may exist, but XML must be used in the same way in

the same application or specification…or not at all.

XML is both machine-readable and human-readable. XML tags are verbose in the sense that words are fully spelled out in the tags, so it is easier to see what is meant. Unlike HTML, XML allows few acronyms and abbreviations. It is possible then to interpret an XML document without special training or a glossary of tags.

A document that declares its own data elements is a standalone instance of XML, but declaring a set of standard, reusable tag names and definitions makes XML more useful. With a document type definition (DTD), many documents can use the same tags by simply referencing the DTD. XML schemas (a form of DTD) are now used for this purpose because in an XML schema, one can specify data types (e.g., string, currency, date), declare namespaces (pointing to a web location for its data element names and definitions), and more. An XML style sheet (e.g., developed following the XML stylesheet language specification [XSL]) can serve as a standard set of instructions on how to structure and display an XML document in human-readable form.

XML Technology

Here XML technology is the term for the broad subject matter covered in Organization for the Advancement of Structured Information Standards (OASIS) at *www.oasis-open.org*.

People rarely give a second's thought to most aspects of paper business transactions since they are so familiar and routine. In the XML environment, however, business is conducted through information exchanges that initiate, describe, and fulfill transactions between parties. XML technology must be able to address all aspects of electronic business transactions.

In going from paper and paper-like electronic documents to fully electronic XML-type records, it should be no surprise that the technology involved is not simple. Structuring the electronic information exchanges, handling security and authentication, defining technical architecture models, dealing with the Internet across which XML exchanges occur, and more all suggest that there is much more to XML than simple data tagging.

Everything done for conventional

The Key Role of XML in the Legal Environment

In areas such as law, the courts systems, safety, and justice, agencies have seen that a future freed from the constraints of paper is desirable. The success of courts and their electronic filings and electronic document management is interrelated with the success of all others courts that would implement such systems. While one political jurisdiction (e.g., a county) might "do XML" alone, there would be little acceptance for it if every law firm or litigant had to master a different data model and maintain separate forms, style sheets, and a different technical architecture for each court in which they had a case.

A court could, for example, build an effective electronic filing and document management system and, in doing so, become but an island of technology. King County in Washington investigated whether there were standards other courts were using or were considering. If each court took its own approach, litigants and law firms could not afford to accommodate the many methods. They would have no choice but to stick with paper documents – the traditional, or default, standard for court records – and to employ familiar methods and accessible technology. No one alone would be able to gain from the fully electronic document enabled by XML technology. There were no standards, so King County set out to help develop them.

"LegalXML" was formed in 1999 by court managers and clerks, vendor representatives, academics, and others who shared a vision of the potential from using XML in courts, law enforcement, and related areas. LegalXML, now a member section in OASIS, assumed the work of building technical standards for XML-based electronic filing in state courts, integrated justice districts, legislation, e-notarization, e-contracts, and related legal subjects. King County and Washington's State Administrative Office of the Courts got involved in the work of LegalXML because it mattered to everyone's long-term success. One motivation was that of automating a great deal of data entry; this promised substantial savings in "keystroke labor" used now to transfer information in "read-and-retype" fashion. There would be little hope for gaining such efficiency without markup standards for routine tasks in processing court documents.

About 50 million pages have been imaged, creating electronic versions of the paper documents in court case files. From imaging, workflow, and electronic document management there is now extensive efficiency in work processes. File users have achieved some freedom, too. There is a world of difference between going to the county clerk's office to check out the one-and-only case file and opening a browser window to access the files from one's desk. Processing XML marked up data and litigants' e-filing XML-based documents are visible through the windshield, and paper is in the rear-view mirror.

records (e.g., task-appropriate formatting, ensuring authenticity) must be done for XML-based records. Whether working with information on paper or with electronic information exchanged through XML technology, the same, or parallel, issues must be addressed to do business safely and successfully. Consider a few examples of what must be handled, both for paper and now for XML type records and information:

1. Information is captured by writing it down, creating a document with details necessary to make a record of a transaction, sometimes using standard forms or style sheets to do so. In XML technology, specifications that define schema and style sheets address this issue.

2. Definitions and standard terms are specified to help us avoid misunderstanding. In XML technology, data models, vocabularies, and namespaces treat this.

3. Exchanges are protected and directed in paper e-records with standard methods of packaging and transmittal that are trusted to be accurate, secure, and timely; that is, there is a standard technical architecture that is trusted to ensure the security and authenticity of transactions. The XML and OASIS specifications relating to messaging, security, and architecture help deal with this.

4. Conventional indicators supply evidence of actual intent, including signatures executed in accepted, trusted ways. Strategies for the equivalent in electronic technology are evolving. XML technology includes digital signatures and other authentication and validation specifications to provide for this need.

5) To ensure that transactions are not electronically hijacked and changed, exchanges are secured – and perhaps encrypted. Significant challenges arise when exchanges take place across the Internet. A substantial number of security-related specifications for XML have been developed to help designers ensure business conducted with XML technology is safe and reliable.

For both paper and XML-based records, records management principles are employed. The tested and trusted methods used with paper-based records may require modification, or different tools for managing all-electronic XML based records may be needed. RIM professionals learning about XML technology need to articulate what, if anything, should be done differently to manage those electronic records.

Roger Winters *is a program and project manager for the Department of Judicial Administration, King County (Seattle, Washington). He is the editor for the Electronic Court Filing Technical Committee in OASIS (Organization for the Advancement of Structured Information Standards). He may be reached at* Roger.Winters@metrokc.gov.