XML Makes Its Mark

Save to myBoK

by Rachael Sokolowski

The advent of managed care and shifting alliances between HMOs, hospitals, and providers has fragmented the industry into separate entities that are highly dependent upon each other for the exchange of patient information. But that information, in the form of a medical record, is on paper—not a form easily processed by computers. Documents, whether on paper or computer, must be in a form in which humans can read and analyze and one in which machines are able to process.

Heading into the Future

There has been a recent leap in Internet usage by the healthcare industry to manage information. However, HyperText Markup Language (HTML), the most popular language for creating Web sites, has many limitations. These limitations become obvious when you try to search for information on the Internet and receive links to irrelevant Web pages and imprecise information.

However, the next generation of the Internet has arrived with the specification of eXtensible Markup Language (XML). Just like HTML, XML can define how pages appear in Web browsers. But XML does much more than that—it provides a context for information. While HTML tells the browser what a page should look like, XML defines what the piece of information actually means. Both HTML and XML are markup languages; that is, they insert markup, or additional information, into text. The markup is known as a "tag." For example, HTML marks up text on a page by inserting a <B> tag when the text should be bolded. But the HTML tags leave no way to distinguish between a patient's name and an item on a purchase order since it tags them both with the <B> for bold. XML, on the other hand, marks each item with more specific tags that enable computers to distinguish them. With XML, humans and computers can tell the difference between a <DIAGNOSIS> tag and <SYMPTOM> or a <Medical.Record.Number> from a <Phone.Number>.

In the past two years, a growing number of providers, vendors, and standards organizations have been investigating how XML can solve problems for the paper and electronic medical records. While XML holds great promise, there are some challenges to be faced before the true potential of XML-based healthcare information is realized.

The "Super Highway" of Information Delivery

With the amount and diversity of documents in paper-based medical records, moving to electronic medical records (EMR) systems will take some time. For instance, one challenge includes the fact that standard descriptions of the types of documents in the paper-based medical record do not exist. To get a better handle on how XML can help, it's best to start at the beginning. What is XML?

First and foremost, XML is a markup language. As previously stated, markup languages are technologies that describe electronic representations of paper documents to a computer by inserting additional information into text. Traditionally encompassing proprietary formatting information, the information expressed in markup languages contains two pieces: the content (text) and the format (visual representation).

As a markup language, HTML has a fixed set of markup, which specifies the formatting of information for display on a computer screen.

```
<HTML><HEAD>
<TITLE>Encounter Registration
</TITLE></HEAD>
```

exhibit 1—HTML prescription format

In the HTML version of the prescription (exhibit 1), it is nearly impossible to accurately determine the medical

```
<BODY bgcolor="#FFFFFF">
<H1>Patient Information</H1>
<P>Jane Doe<BR>
MRN:123456789<BR>
DOB: May 13, 1923<BR>
Address:123 Main St., Anytown, USA (home)<BR>
Phone:555-345-9876 (home)
```

record number for the patient. The tags <H1> <P> and
 are instructions to the web browser to display the information, where:

- <H1> indicates a first level heading
- <P> signifies the beginning of a paragraph
-
 denotes a line break or carriage return

However, none of these tags helps the user understand what the information contained in the page is or how to search for the medical record number.

While XML is a standard markup language, it separates the content of the electronic document from the format or the visual presentation. It is a language that allows one computer system to describe to another (in a standardized way) the markup and structure used within a document. In healthcare, XML codes such as <Prescription> and <Refill> can easily identify important or meaningful information. The following text represents the content of a prescription:

exhibit 2—content of a prescription

```
Amoxicillin, 250 mg. Capsule, 30, 1 cap(s), 3 times daily until gone, no refills, can substitute generic equivalents
```

A structured prescription represented in XML might be represented as follows:

exhibit 3—XML prescription format

```
<Prescription>
<Medication.Name>Amoxicillin</Medication.Name>
<Form>250 mg. Capsule</Form>
<Dispense>30</Dispense>
<Dosage Amount="1">1 cap(s) </Dosage>
<Instructions>3 times daily until gone</Instructions>
<Refill Number="0">no refills</Refill>
<Substitute>can substitute generic equivalent</Substitute>
</Prescription>
```

Information expressed in XML can be identified and located by the tags, which describe content. In the textual description of the prescription, it is nearly impossible for a computer program to determine the number of refills. In the XML representation, it is an automatic task for a computer program to extract the exact number of refills:

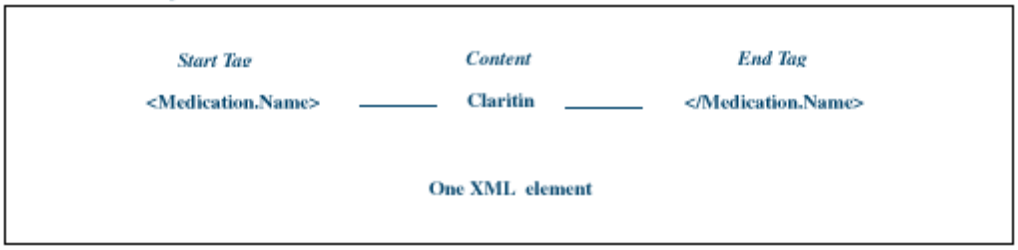
exhibit 4—refills

TEXT	XML
No refills	<Refill NUMBER="0">No refills</Refill>

In XML, the tags are called elements—named units of information that can be defined.

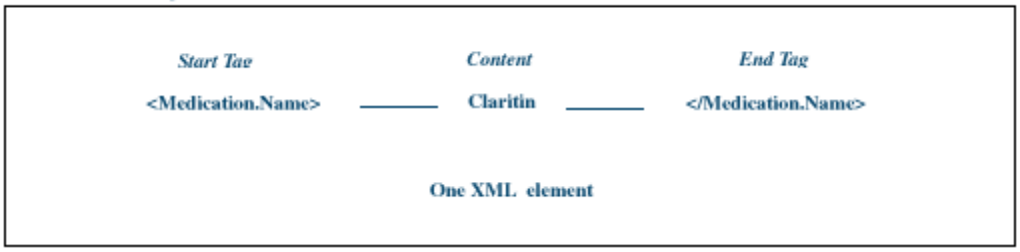
In the next example (exhibit 5), the name of the tag is "MEDICATION_NAME." Elements have boundaries, e.g., they specify a beginning (the start tag) and an ending (the end tag). Between the tags lies the content—in this case, the name of the allergy drug, Claritin. Further information that describes an element is called an attribute. Attributes, immediately specified after the element, have potential application in healthcare since they can represent alternative forms or representations of words, such as codes.

exhibit 5—tags and element



XML allows any number of attributes on a given element to be defined. In the example in exhibit 6, the name of the tag is "Immunization" and the name of the attribute is "CPT" to represent the coding system. In addition, attributes possess values. In this example, the value for the CPT code attribute is 90724 (the CPT-4 code for an influenza immunization).

exhibit 5—tags and element



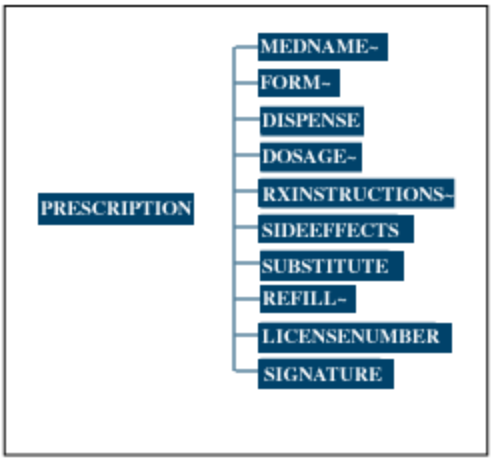
The names of the tags and the rules for using them are contained in the document type definition (DTD). The DTD describes the structure of the document and defines the names of tags it contains. It also declares the order in which the tags occur and how often the tags can appear (e.g., the DTD defines the hierarchy of the tags). A DTD for a prescription might contain structural elements for the medication prescribed, the dosage, the form, and the quantity.

XML DTDs are important for healthcare informatics because they will provide context for narrative text, a document information model, agreement on high level structures, and a facility for standardizing formats (or the visual presentation with style sheets).

In the example in exhibit 7, a prescription tag contains 10 sections:

- medication name
- form
- dispense
- dosage
- instructions
- side effects
- substitutes
- refills
- license number
- signature

exhibit 7—graphical representation of an XML DTD of a prescription



The DTD will specify whether the order or the sequence of the tags is important. In this example, for a prescription document, the order is important (e.g., the medication name is followed by the form and the form of the medication must be followed by the dispense amount).

As shown, XML has great potential for representing paper-based documents in an electronic form, once the paper documents have common and identifiable structures. Again, deriving such an electronic representation requires standard forms for different document types. It also requires a mechanism for locating, storing, and querying the electronic document, as well as the potential for a DTD in some type of patient record repository. Currently there is no standard set of document types for healthcare, nor is there a commonly defined service for accessing the XML documents. However, to address this issue, the American Society for Testing and Materials (ASTM) E31.25 subcommittee is developing a standard set of DTDs for healthcare documents.

XML DTDs for Health Care

The ASTM E31.25 is a subcommittee of ASTM E31, the committee on healthcare informatics. The intent of the subcommittee, named XML DTDs for Health Care, is to develop standard electronic document representations of paper-based healthcare documents and forms. A goal of the subcommittee is to work together to enhance existing levels of interoperability among the various XML/SGML standardization efforts (which include efforts within HL7), products, and systems in healthcare.

The scope of E31.25 is the development of standards and promotion of knowledge related to DTDs for healthcare. The subcommittee is developing implementation guides, sample document instances (versions of the document with XML markup), and a validation facility for verifying conformance to the voluntary ASTM standard DTDs. The DTDs are being developed under the current XML standards and will adapt as the XML standard evolves.

The first priority for E31.25 is to develop standard document type definitions for transcribed documents. To reach this goal, the subcommittee coordinates with other ASTM subcommittees and outside organizations with related interests and standards, for instance, HL7. HL7 is using XML as the syntax for HL7 messages to request and send healthcare information. In addition, the HL7 patient record architecture is being developed as an exchange mechanism for electronic documents.

The focus of ASTM E31.25 is document-centric. This focus differentiates activities from other ongoing efforts that focus on:

- object-oriented technologies (ASTM, CEN, CORBAmed, HL7)
- healthcare messages and transactions (HL7, X12)
- document exchange (HL7 patient record architecture)
- financial transactions (XML-EDI, X12)

E31.25 views the medical records as a collection of electronic documents and incorporates other efforts when possible. A significant work in progress is document analysis. The document analysis process answers questions such as:

- What kinds of documents exist, and what common types of documents can be identified?
- What are the basic structural and logical components that occur within each document type?
- In addition to text, what other types of information exist and in what form (e.g., x-rays, dictations)?
- What are the logical relationships between each of the items in the document?

Conclusion

XML has great potential for representing the paper documents in the medical record as well as forms and other paper-based documentation. A standard set of document types within the medical record must be developed to move the information to a digital form. Once the set of XML document types are developed, there will be a common electronic representation. In order for XML to become the solution, however, there must be coordination between standardization efforts so that the document descriptions, tag names, and attribute representations do not diverge.

Rachael Sokolowski *is president of Magnolia Technologies in Arlington, MA, vice president of research at iTrust, co-chair, HL7 XML/SGML special interest group, and co-chair, ASTM E31.25, XML DTDs for Health Care. Her e-mail address is rsokolowski@magnoliatech.com.*

Article citation:
Sokolowski, Rachael. "XML Makes Its Mark." *Journal of AHIMA* 70, no.10 (1999): 21-24.
