THE MANY FACES OF

Peek under the hood of almost any application today, and you'll likely find some aspect of XML technology hard at work.

BY KELLI WISETH

Before the advent of Extensible Markup Language (XML), enterprises that wanted to integrate either systems or data had just a few choices. They could implement gateways to support interactions among disparate systems. They could create central data or message hubs, into which all communications would be converted into the appropriate languages of other participants. Or, they could implement some combination of both these approaches.

The problem was, all these approaches very likely had the word "proprietary" buried somewhere at their core, and adding more than a handful of companies, trading partners, or technologies into the mix only added to the complexity—and the cost. When it came to trying to integrate systems or data, it definitely wasn't a case of "the more the merrier," and harnessing the full potential of the internet always seemed just out of reach. The architectural diagrams associated with many integrations looked like paintings by Jackson Pollock.

With XML, however, much seems to have become possible, even things that may not have been envisioned when the specification first emerged. In a basic sense, XML is meeting an enormous need in the information infrastructure by providing the rules to create a standards-based lingua franca among disparate systems, according to Susan Feldman, research vice president for Content Technologies at IDC. For example, just as people who speak only a single language—be it French or Swahili—can communicate only with speakers of the same language, says Feldman, "as soon as you have a language that goes across those groups, then you have people interacting from different cultures. That's really what XML is bringing to applications—if the applications can take XML in and send XML out, they can communicate."

XML not only facilitates communication but also:

Enables entire industries to forge supply chain links, exchange information, and integrate processes as well as data.

Allows consumers to access bank accounts, order new services, look up driving directions, and get all manner of other information. The richness, format, and amount of content that gets delivered to the device can vary, depending on the bandwidth and display capabilities of the target device (via broadband from home, versus dial-up from a hotel room while on the road, versus via wireless PDA or cell phone, and so on).

Provides the foundation for new technologies and protocols, such as Web services (embedded as the payload format in SOAP packets, used to describe a Web service's capability, using the XML-based Web Services Description Language, or WSDL, and so on), business-process management, enterprise-application integration, business-to-business interactions, and application-to-application interactions built on open standards languages

created using human-readable and machine-readable text.
Forms the basis for messages and events that get shared throughout the Oracle technology stack, enabling Oracle Database, Oracle Application Server, Oracle Collaboration Suite, and Oracle E-Business Suite applications to integrate transparently, not only with Oracle technologies but also with components from other systems, such as SAP, WebLogic, and so on.
Is used to create industry-specific dialects that enable message creation to support interactions among systems.
Supports publishing or content-management systems and provides the means to easily create, store, manage, retrieve, search, import, export, and repurpose the content that makes up books, magazines, legal publications, and technical manuals, for example.

Any application that needs to transmit information, content, or data, or to interact in processing scenarios with other applications, is most likely using XML-based technology to do so. All these scenarios and more continue to emerge, thanks, in part, to XML's chief characteristic—extensibility.

X IS FOR EXTENSIBLE

XML is a self-describing data format. "Self-describing" means that metadata about the content goes along with the content itself. That is, XML documents (or files containing XML markup) include information in the file that conveys to the recipient—human or machine—how to interpret the tagged content and structure of XML.

XML emerged in the late 1990s in response to the need to extend HyperText Markup Language (HTML). As with HTML, XML is also derived from SGML (Standard Generalized Markup Language), which has been an ISO standard since 1986 and is used widely in the publishing world. Businesses such as publishing houses, government agencies, and defense contractors have long used SGML-based systems to manage large-volume documents—a repair manual for a passenger jet that details every nut and bolt, for instance—"to structure documents, to make sure that all the right sections are in the documents," explains Rita Knox, Gartner Inc. analyst.

Unlike HTML, which provides the strict set of tags that render Web pages in a browser, "XML provides the basic alphabet," says Knox, "with which to create new markup languages."

XML can be thought of as a "meta-markup language," adds Mark Scardina, Oracle group product manager and XML evangelist, "because it enables users to define their own markup languages to describe and encapsulate data into XML files." Using the rules and conventions described in the XML recommendation (or the XML Schema recommendation [www.w3c.org/TR/#Recommendations]), anyone can sit down and create his or her own XML dialect and define the structure as well as create the rules and conventions with which to then mark up documents for transmission, display, or submission to another system.

XML is being used to create new languages for health, finance, insurance, taxation, government, and numerous other industries around the world to perform a wide range of tasks. Some examples include:
Geography Markup Language (GML), an XML-based encoding standard for the transport and storage of geographic information, such as the geometry and properties of geographic features.
Scalable Vector Graphics (SVG), an XML language standard approved by the W3C (www.w3.org/TR/SVG/) for describing two-dimensional graphics in XML, specifically vector graphic shapes, such as paths consisting of straight lines and curves, images, and text.
papiNet (www.papinet.org), an XML-based processing standard for the paper and forest supply chain.

In addition to being the basis of new languages, XML underlies the burgeoning Web services application development and deployment model.

X IS FOR EXTENDING WEB SERVICES

Web services are beginning to deliver on the vision of interoperable, distributed computing that has been eluding the industry since initials like "DCE," "CORBA," and "DCOM" were first strung together (that is, starting in about the late 1980s). That vision is becoming reality thanks in large measure to the evolving suite of XML recommendations—XSLT, XML Namespace, and XML Schema, for example—that provide the platform- and programming-language-neutral data and communications mechanisms necessary for such interoperability among disparate platforms, operating systems, and programming languages. These mechanisms include the following core Web services protocols:

Simple Object Access Protocol (SOAP), an XML-based protocol that uses a standard character encoding (which means it can be processed on any platform). In simple terms, SOAP carries XML-based messages (information and instructions) between Web services.

Web Services Description Language (WSDL), an XML vocabulary for describing a Web service and its specific capabilities.

Universal description, discovery, and integration (UDDI) defines XML-based rules for directories (registries) in which Web services can be located and in which companies can advertise their Web services.

Just what is a Web service, though? According to the W3C's Web Services Architecture working draft document, "a Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards." (Visit www.w3.org/TR/2003/WD-ws-arch-20030808/.)

In simple terms, an example of a Web service in action might be something like this: A developer builds a credit-card-verification service using the language and tools of his or her choice—C, C++, C#, or Java, say, running on a J2EE application server. A description of the service and how to communicate with it—what inputs and outputs the service can accept, and in what format and structure—is described using WSDL, the XML vocabulary developers use to describe the Web services they create. Information about the Web service, who's hosting it, and where to obtain the WSDL is published in a UDDI registry.

The basic Web services protocols (SOAP, WSDL, UDDI, and XML) are just the beginning, however; the direction of Web services technology is toward enabling interoperability

at a higher, larger-grained level, that is, at the business process level, across multiple businesses and even across industries. And XML is at the core of many of these initiatives, directly and indirectly. For example, the emerging Business Process Execution Language (BPEL), which relies on WSDL and WSDL extensions, is just one of the many emerging standards for dealing with the notion of long-running distributed business processes (rather than simply distributed data). Further work in the area of Web services transactions, choreography, and other XML-based specifications and protocols will ultimately lead to enterprise application integration and business process scenarios that shorten supply chains; save time; leverage existing investments; and achieve more-streamlined, more-effective interactions among businesses.

X IS FOR EXPANDING UNIVERSE OF APPLICATIONS

Oracle is leveraging the power of XML technologies across its product lines in various ways. That means, for example, that a purchase order being sent over the Web, via Oracle Application Server (using XML to stream messages, using the Integration component), can be leveraged by the Oracle E-Business applications, and vice versa.

At another level, much of the configuration and setup information to do with installation or configuration of Oracle products uses XML. For example, Oracle JDeveloper 10g and its new Application Development Framework (ADF) essentially adapts much of the user interface components and makes menu options available based on the technology options to your application technologies. This is in part because XML metadata is being used under the covers as you configure the layers of the Oracle ADF framework, for example, to produce the declarative options for development while accommodating custom coding wherever necessary.

Oracle's use of XML throughout the product stack also means that Oracle products can be integrated with or leveraged by other, non-Oracle-based systems more easily. For example, Oracle E-Business Suite applications use XML for message exchange among modules. Since Oracle XML Gateway supports all document type definitions (DTD)-based XML standards, Oracle E-Business applications can integrate with external systems that support any of the published DTD that have emerged from various standards bodies, including OAG, RosettaNet, and iFX. As another example, Oracle Calendar, a component of Oracle Collaboration Suite, relies on XML and Web services to enable applications to retrieve, through common XML queries, calendaring data for display in any portal, client application, or back-end server.

Not only Oracle but organizations everywhere are harnessing the power of XML in various ways, from simply incorpo-

rating XML techniques in the user-presentation layer to developing leading-edge business platforms for extensive business process and supply chain integrations.

For example, an e-marketplace in Brazil is bringing technology, much of which relies on XML, to bear on streamlining and automating the pulp and paper supply chain.

K IS FOR FLEXIBLE INTEGRATION

Based in São Paulo, Pakprint is the leading network for the Latin American pulp and paper industry formed by International Paper, Votorantim Celulose e Papel (VCP), Suzano, Bahia Sul, Klabin, and Ripasa. Pakprint was developed for the benefit of the entire industry and today represents 90 percent of the printing and writing business in Brazil and a network of 7,000 suppliers and 1,000 direct customers. More than 40 percent of the paper industry's order process follows the made-to-order process, according to Pakprint CTO Tânia Datti, "which is much more involved and time-consuming than the stock-to-order process." That means that the time from initial order to delivery can vary widely, depending on the product characteristics. Additionally, the orders traditionally have been conducted by fax machine and require repetitive manual data entry.

To gain better visibility across the supply chain, improve distribution, and streamline trade mechanisms, Pakprint integrated the ERP systems of the five-company consortium and hundreds of buyers, including a wide variety of companies from within the paper industry: manufacturers, dealers, graphic service providers, publishing houses, and suppliers.

The system—built using a variety of Oracle products, including Oracle9i Application Server, Oracle E-Business Suite, and Oracle iStore—has eliminated the time that people in the industry spend to type orders, receive and send faxes, retype data, and work out details on the phone. Instead, buyers now have several options, including

using the portal Web site—each participating company has its own storefront within the portal, along with associated trading rules. The storefronts are integrated with each company's ERP system, providing such services as pricing inquiry, backorder status check, industry news, and a showroom area for partners and customers.

Elements such as purchases are defined as XML files, and DTD files are generated from each company's mapping definitions to create common views and guarantee complete compatibility with the XML files sent and received through the system.

The portal Web site is just one way to buy paper using the Pakprint system, according to Emerson Iezzi and Felipe Dias, systems engineers for Pakprint. There's also a Java client application, Postman, which lets those buyers with ERP or other systems send their purchase orders—they can generate XML files or even spreadsheet or text files. "All orders received from buyers (using Postman or any other enterprise application integration [EAI] software) are processed and sent to the appropriate target industry company," explains Dias. PakPrint's Datti adds that the system now has the flexibility to accept orders from any client. "It doesn't matter if orders come from the Web site, from EDI, from system to system, XML, or mobile," says Datti, "now it's fully integrated."

The business rules driving any specific integration between buyer and producer can also be configured with a high degree of granularity. For example, "using XML files, an integration between buyer and a company's back end can be set up so that the buyer can actually choose the plant that produces the paper," says Datti, "so the integrations are very specific to the business needs of the individual companies, industry type, and buyers." Datti explains, "XML gives us a lot of flexibility to make the data transformations required for our integrated business partners. Using XML gives us a faster way to adopt documents and versions to achieve a specific business need from customers and buyers."

Since going live in 2002 with just a handful of companies, Pakprint's portal has grown to integrate 170 customers and 70

buyers and each day processes some 10,000 transactions for trading partners using the Pakprint portal.

Although the system is based on Oracle products, network members, trading partners, and buyers can use any type of system to work with the system because it relies on XML, including an XML standard for the paper and cellulose industry, papiNet. According to Datti, "papiNet encompasses a broad spectrum of business processes for the pulp and paper industries, based on best practices, created by a global industry network."

"Currently, Pakprint uses only the XML message formats of papiNet, but since papiNet is a business process standard, using papiNet will enable us to implement full business process management when the time is right for our network member companies," says Datti.

"It was critical that we had an easy way for participating companies to integrate with the community, and Oracle9iAS Integration, with its prebuilt adapters and strong XML support, provided a robust architecture," Datti adds.

In addition to XML being a key enabler to integration scenarios such as this, it's also being used to solve completely different types of problems and meet vastly different needs. For example, Iceland's eMR Consulting created an important questionnaire application for deCODE genetics initially using the Oracle XDK (see "Oracle XML Developer's Kit 10g") and has since migrated the application to Oracle XML Database (XML DB) to take advantage of native XML features, such as XML Schema support, XPath functions, and tight integration of XML with both SQL queries and Oracle Text indexing.

X IS FOR XML DATABASE

Headquartered in Reykjavík, Iceland, with offices in Boston, Chicago, and Seattle in the U.S., deCODE genetics Inc. is a medical genetics research lab that focuses on population-based genomics, supported by state-of-the-art genotyping facilities, bio-informatics, and medical-informatics research for mapping disease-causing genes.

Oracle XML Developer's Kit 10*g*

Oracle has released various versions of its XDK—XML Developer's Kit—since Oracle8*i*. With Oracle XDK 10*g*, the entire set of XML libraries and components—for Java, C, and C++—that previously were released separately have now been packaged and included as a single XML Developer's Kit with corresponding components and libraries.

The XDK libraries can be used in the database, on the middle tier, or in the client, which means you have flexibility and choice when it comes to architecture and deployment, says Mark Scardina, group product manager and XML evangelist for Oracle.

Furthermore, in the new release the functionality "has been filled out so that building applications against the database is a lot easier and more efficient," says Scardina. For example:

XDK support for XMLType. Oracle has introduced new XML C and C++ APIs that support both file-based and direct XMLType DOM operations. "This greatly simplifies building XML-enabled applications against Oracle XML DB," says Scardina, "because now developers can use the same XDK DOM interfaces for directly accessing XMLType DOMs in the server, thus improving performance of their OCI applications as compared to previous releases because they have eliminated a set of serialization and reparsing steps."

XSLT 2.0 support. XSLT 2.0 introduces datatypes into the transformation process. "This is particularly important from a database perspective," says Scardina, "because now you can do transformations not simply of the syntax of XML, but you can perform data operations and datatype validations on the XML." The fact that the

Oracle database can store XML natively and do transformations that take into account these datatypes "makes a more powerful combination for application developers."

Full internationalization support. XML is based on Unicode, which means it supports all possible character sets. Oracle Database 10*g* is now a fully Unicode-capable database, and the Oracle XML Developer's Kit (XDK) libraries are fully integrated with Oracle Database NLS (national language support) and globalization libraries, enabling developers to work seamlessly in dozens of languages and encodings, says Scardina, "which becomes extremely important outside of the United States—especially in Asian or Middle Eastern countries, where you've got multibyte character sets with very unusual encoding. The Oracle XDK and Oracle Database can fully support those character set encodings."

JAXB Support. The Oracle XDK 10*g* now supports the Java Architectural Binding for XML (JAXB) specification that provides a Java object access to XML documents. "JAXB is an important XML Java extension for Java developers," says Scardina, "as it provides a very easy-to-use set of interfaces when working with XML. With JAXB, source code can be generated directly from XML schemas to bootstrap applications that need to create or access XML documents, thus insulating developers from the arduous task of directly working with the XML structures."

These are just a few of Oracle XDK 10*g*'s new features. For full details, visit Oracle Technology Network. XML Developer's Kit ships with Oracle Database 10*g* and Oracle Application Server 10*g* and is also downloadable from the XML Technology Center of OTN (otn.oracle.com/tech/xml).

"**W**e have been very pleased with our experience of the XML support in Oracle. Oracle XML DB has enabled us to develop our mission-critical questionnaire system, called Questor, with less effort and complexity than otherwise possible using traditional, nonnative XML techniques."

—Hákon Gudbjartsson, Vice President of Informatics, deCODE genetics

Population-based research involves studying large groups of people to learn how genes affect health and involves lots of data gathering using surveying techniques with study participants. Approximately 50 ongoing research projects at deCODE require supporting surveys, each of which has hundreds of questions (typically 250-300) and thousands of participants (10,000 to 20,000).

To facilitate conducting surveys electronically, over the Web, rather than continuing to use error-prone and expensive paper-based techniques, deCODE (with consulting help from eMR Consulting) designed and developed a survey system, called Questor, centered around an Oracle Database.

The functionality of XML DB is central to the system and its capabilities, especially its ability to handle XML natively. XML DB is a core capability of Oracle Database, Release 9.2 and higher. It's installed automatically during installation of either the standard or enterprise edition of the database. XML DB leverages the object-relational capabilities of Oracle to support XML natively. "We have been very pleased with our experience of the XML support in Oracle. Oracle XML DB has enabled us to develop our mission-critical questionnaire system, called Questor, with less effort and complexity than otherwise possible using traditional, nonnative XML techniques," says Hákon Gudbjartsson, Ph.D., vice president of Informatics, deCODE genetics.

According to Magnús Kristjánsson, an eMR Consulting developer, the system was built to meet three key design goals: to be able to support multiple authors composing questions and surveys; to support multiple presentation channels, including browser and PDA; and to be able to integrate with multiple subscriber and surveyor systems.

The system developers created a survey markup language for exclusive use with this system, defined in XML, using XML Schema, to handle transmissions of all data to and from the system. The survey markup language governs the structure of the surveys and is used for the content— the actual questions that comprise the surveys and the participants' answers, says Kristjánsson. He points out that Questor could be used for all manner of polls, exams, or surveys, not just questionnaires for bioresearch, and that deCODE creates the actual questions—an art in itself.

The surveys are in XML. Each survey has links to questions, which are stored in the repository, also in XML format. Developing a repository of good, proven, consistent questions was an important goal in developing the system, says Kristjánsson: "It's an art of its own to write a good question, a question about which there can be no doubt as to the meaning of the question when participants give answers." Currently the system contains about 4,000 reusable questions.

"The Oracle Text XML search facilities, combined with our topic hierarchy, also defined in XML, enable us to quickly locate relevant and reusable questions when creating new surveys. Storing participant answers in XML format enables us to easily transform the survey results and import them into other research systems for further analysis, in particular our Clinical Genome Miner system," says Gudbjartsson.

Oracle Database 10g's XML DB serves as the repository. "The XML Repository feature supports our document-oriented view of an XML-based repository of reusable questions that can be deployed repeatedly and consistently across a number of surveys," Gudbjartsson remarks. "Overall, at deCODE, we are very pleased with the excellent work eMR's consultants have done on the Questor system."

X IS FOR THE UNKNOWN—WHAT'S NEXT?
Many emerging technology initiatives have at their core a dependence on the 32-page XML recommendation from the W3C and its complementary recommendations and initiatives that continue to evolve.

While the ancestral markup language from which XML is derived, SGML, was never intended to be used to exchange information among multiple, different computing systems, the design of XML, on the other hand, is expressly made to be extended. So far, XML dialects, protocols, and applications are enabling integration of data and processes. The next frontier may well be agent technologies built on XML and ancillary XML-derived standards that might begin to deliver on some of Tim Berners-Lee's notions about the "semantic Web." ∎