

Technology Hot Spots

XML in Printing and Publishing

By James E. Harvey, President of Media4theWorld, LLC

If you're like most printers and prepress professionals, it suddenly may seem that, everywhere the extensible markup language (XML) is being touted...in graphic arts systems brochures, press releases, on the floor of tradeshow, and in news of developing standards. Is this just another buzzword where suddenly everyone's got one and systems that you thought you knew claim to have greater scope than you ever imagined, yet little has changed and nobody is quite sure of the buzzword's "true" definition? And what exactly is XML, and why should you care?

The Roots of Markup Languages and XML

First of all, XML's roots are print production. And like anything else, if you know its history you'll have a better understanding of what it does, how it does it, and why it does what it does.

In the late 1960s electronic phototypesetting displaced conventional hot-lead typesetting and publishing systems became computerized. However, early typesetting and formatting commands were embedded directly into text files, which posed a problem for publishers—they could not transfer work from one typesetting service to another because most commands were proprietary and often constituted 100%-plus overhead in their text files. Often moving a job from one service to another necessitated the cost of re-keying work.¹ Publishers began to demand that generic markup be applied to their files and that the markup be replaced with commands through computerized substitution upon output to an imager.

However, even the use of generic markup languages proved not be so "standard" and

could be difficult to interchange. The computer syntax, notation, and " delimiters could include the use of characters such as "<", "[", "*" and others. However, many systems used the same delimiters for different functions, so substitution and conversion from one generic markup language to another wasn't so simple.

Enter SGML

In the mid-1970s the Graphic Computer Communications Association (later renamed Graphic Communications Association and recently renamed again IDEALiance) formed a committee to study generic markup languages and to create a standard markup language. The committee based their standard on the Generic Markup Language developed by Charles Goldfarb at IBM. Goldfarb was a lawyer rather than a programmer. In 1968 he was asked by IBM management to interview legal publishing firms to determine why they were unhappy with their legal publishing and information systems. Goldfarb determined that the systems' functionality was good, but the basic problem was that each of IBM's systems required a different text notation and markup.

The result of GCA's work was Gencode®. Gencode® was not a set markup scheme or a set of predefined tags for publisher's use; rather, the committee settled on a descriptive computer language that users could employ to generate their own markup schema.² In 1983 Gencode® became an ISO work item that was approved in 1985 and was published in 1986 as ISO 8879 under the name Standard Generalized Markup Language (SGML). Every SGML application has three components:

1. The declaration

Where Do I Get Parsers?

- SP (A free, object-oriented toolkit for SGML parsing and entity management available for Windows, Unix, and Linux) <http://www.jclark.com/sp/index.htm>
- ARCSGML 1.0 (Charles Goldfarb's Almaden Research Center SGML Parser — available for Windows and Macintosh) <ftp://ftp.ifi.uio.no/pub/SGML/ARC-SGML/>
- YASP (Pierre Richard's Yorktown Advanced SGML Parser or "Yet Another SGML Parser" available for Windows and NT) <ftp://ftp.edf.fr/pub/SGML/YASP/>
- MSXML (Microsoft XML Parsers. Microsoft offers several parsers with different features) See: <http://support.microsoft.com/support/kb/articles/Q278/9/69.ASP>
- XP (A high-performance XML parser in Java, as well as other parser tools and test kits) <http://www.jclark.com/xml/>
- XML Parser for Java (IBM's Alphaworks available for Linux and all Java platforms — Provides W3C schema validation.) <http://www.alphaworks.ibm.com/tech/xml4j>
- XSV (an open source XML Schema Validator from University of Edinburgh/W3C) <http://lists.w3.org/Archives/Public/xmlschema-dev/2001Jul/0067.html>
- Xerces (XML Parsers in Java and C++ from the Apache XML project, lots of great stuff for developers, provides schema validation support) <http://xml.apache.org/>

Also, check out XML Spy's integrated development environment and suite. It includes embedded validation tools and is used by most of the graphics standards developers that I know of: <http://www.xmlspy.com/>

Tip: Not all parsers perform equally. As part of your IT operating procedures, it may make sense to select tools that support or embed a common browser. (Think RIPs!)

2. The Document Type Definition (DTD)

3. The document instance

In SGML the “declaration” is used to define the syntax that will be used for the user’s markup application in a way that is machine-readable and that can be processed. The declaration sets the character set to be used, special data types that will be used such as graphic file formats, sets limitations on the name length of any element, sets the delimiters, and turns on or off a series of optional features that SGML standardized, such as the ability to omit end tags when a subsequent start tag is required so by rule. Declarations can be rather complex.

The DTD is used to define document elements, attributes of elements, and entities, as well as their relationship to one another. A document instance is a text file in which the markup defined in a DTD is applied. A DTD can be used to “parse” an instance and verify its integrity.

Let me stop here and try English. The DTD establishes the rules for how a marked-up document or data file (known as an “instance”) is to be structured and how the tagging is to be applied to the document. The rules may be as simple as “paragraphs are always subordinate to book or article titles.” There are software tools called “parsers” that can be loaded with SGML and XML DTDs or XML schemas (we’ll discuss schema later) and then use the DTD or schema to check that files meet the rules as set in the DTD. (Think preflighting!)

Parseability is the most important feature of SGML and XML schema for data exchange applications. A parser embedded into editors or graphic arts systems can be used to relate commands and queries to the appropriate data elements within a document instance or can be used to relate data elements to their place in a database or other type of record storage. Parsers can be used as standalone applications, but often they

are buried into SGML and XML editors, document database engines, information asset management system, pagination systems, and so forth so that you may not be even aware that they are there. Parsers play a key role in publishing systems that use markup language applications, and you will see that they will play a crucial role in the future of E-business systems in the graphic arts. DTDs and document instances created with the DTDs can also be used to benchmark and test systems that claim to support public SGML or XML standards that are based on DTDs (or schemas...read on)—empowering the user or buyer to test



first-hand whether a system that claims to be “SGML-compliant” is in fact as open as advertised.

It is also important to note that SGML completely divorced formats and fonts from the structure of the content. Early DTD defined “headers,” or “abstracts,” or “paragraphs” as elements and attributes that could include data elements that are not rendered in output but rather provide handles for data management. Examples could include “approval,” “ownership,” or “version,” but adding formatting instructions is generally forbidden.

The earliest adopters of SGML tended to be users whose information had a long life expectancy, such as airplane manufacturers or U.S. federal agencies such as the Treasury Department or the Department of

Defense...organizations that had equipment whose supporting documentation was required to last decades.³ The motivation was application independence, not just independence from formatting instructions. These same companies were pioneering early optical media (what CD-ROMs were called before ISO 9660) and as a result, they made the first use of SGML documents for electronic publishing. In turn, the use of SGML in electronic media immediately led to the development of extensions to SGML that could facilitate media other than text and graphics. In 1992, *ISO/IEC 10744:1992. Information Technology-Hypermedia/Time-based Structuring Language (HyTime)* was published. “HyTime provides standardized means of expressing (1) intra- and extra-document locations, and arbitrary links between them, (2) the scheduling of multimedia objects in ‘finite coordinate spaces,’ and (3) rendering instructions for arbitrarily projecting such objects onto other finite coordinate spaces, and other constructs.”¹ This concept, developed in 1989–1991 by Steve Newcomb, a music professor at Tallahassee State University in Florida, led others to re-think SGML DTD designs.

Document constructs such as text and graphics are variable in length and can’t be easily managed with relational database technologies. Inevitably SGML professionals began exploring object-oriented programming and database techniques. In quick succession in 1990 and 1991, Diane Kennedy of Datalogics and the ATA DTD committee decided to publish DTD fragments or “bricks” that users could build complete DTDs with, rather than complete DTDs that were not good fits for airplane documentation applications. Pushpa Merchant, an XML consultant, developed the brick concept into one that dealt with “frames” of information that were to be self-contained, a concept that is today common in Internet publishing. Finally, Jim Harvey (me-yes, I’ve been called a nerd, but I deny it...more of a “neo-nerd”), then of Volt Information Sciences and the Society

of Automotive Engineers, shed the document roots of SGML altogether and created the first truly object-oriented SGML standard. Rather than defining the ultimate automotive service and diagnostics book structure (which the manufacturers would have never agreed upon), the group based their SGML on objects such as “car,” “assembly,” “component,” and “diagnostics.”

One interesting aspect of the J2008 recommended practice is that it encompasses only the information and the structure of information relative to itself. Although the Data Model is not specific to any data management technique, companies that provide support to OEMs, such as Volt and Datalogics, have begun to model J2008 data in an object-oriented environment that can facilitate these complex relationships.”ⁱⁱ

Around the same time another change was happening in the SGML community. Bell Atlantic Engineers, in 1987, introduced an online service that featured graphic representations of office documents, in color, exchanged over the Internet. They had two options: employ a simplified generic SGML DTD as their exchange format or use the editorial-based IMI format. They picked the wrong option...perhaps one of the worst business decisions ever made! Another product, designed for optical media publishing, called Guide from Owl, Ltd., introduced a simple four-tag SGML DTD that could be used to interpret any document into their retrieval program. Although neither of these first simplified SGML applications survived, some folks at CERN were paying attention, wrote their own simplified tag set, the hypertext markup language (HTML); developed a browser, and gave it away! It caught on and the World Wide Web was born.

The Internet Boom or Bust: E-Business Round One

Tim Berners-Lee, who was then working for CERN, developed HTML in the late 1980s. (Berners-Lee is now the chairman of the World Wide Web Consortium.) HTML uses SGML syntax in a simple set of tags that a Web browser could use to interpret and pre-

sent information. The browser provides the look of the page; style is not carried in the tagged document. (Admittedly, many Web designers use HTML features to jerry-rig their layout.) Furthermore, a simple DOS or ASCII editing tool could be used to create HTML files. The original browser was a free-ware software package called Mosaic. Mosaic is the basis that Microsoft Explorer and Netscape Navigator are built upon. <HTML> is the tag that begins the html document and tells the browser that it is an HTML document. The tag <BODY> begins the part of the document that will display to the screen. Tags that appear between <HTML> and <BODY> are typically meta tags that provide information such as keywords for searchability. This is where the concepts behind XML began to creep into the Internet culture.

The first HTML users were for the most part from the academic community. They didn't care much about formatting or graphics. In 1994 the World Wide Web began to be viewed as something that could be used in the commercial arena, and the need for better presentation and graphics became important. Furthermore, Web tools such as HotMetal and FrontPage began to appear that provided a graphical editing environment. These packages made heavy use of font calls and presentation control.

However, HTML lacked depth and flexibility, and in 1996 the browser wars led to a greater need for additional style functionality; hence, style formats such as Cascading Style Sheets (CSS), and later the XML style language (XLS), were created. Cascading Style Sheets assign design to individual html tags. An example of a large blue italic header might look like:

HI {color:blue; font-size: 18pt; font-style: italic}

HTML caught on in a big way, but soon users were dissatisfied with its limited capabilities and its dumb, document-oriented tag selection. For instance, HTML was used to present products in catalog form on the Web, but once the product and pricing data was moved to HTML for Web processing, there wasn't much

News and Other Good Stuff

Imation Corp. announced its new fully integrated proofing system, the Imation Matchprint Inkjet Proofing System 5042/5060. Optimized for proofing, the system is an eight-page, wide format, roll-fed, six-color inkjet proofing solution. The system is said to be one of the fastest production proofers currently on the market and delivers a full four-page, color-accurate proof in approximately eight minutes.

The Imation Matchprint Inkjet Proofing System 5042/5060 incorporates Hewlett-Packard's thermal inkjet technology with Imation's proprietary Matchprint Color RIP software enabling simulation of a wide range of color hard copy targets, Matchprint watermarked media, and new Matchprint inks.

Key system elements include Matchprint Color RIP incorporating the proprietary Imation Color Fidelity Module (CFM) used in many hardware systems, Matchprint Inks designed for proofing and which come in 680-milliliter cartridges, and Matchprint Inkjet paper to ensure proofing consistency. Need to know more details? www.imation.com.

Kodak Polychrome Graphics introduced its SWORD Thermal Printing Plate, a no-heat, no-postbake thermal plate that boasts run lengths of 400,000-plus. SWORD thermal plates employ a “write-the-background” technology with a negative processing system that is compatible with infrared (IR) laser sources used in thermal platesetters. It has been accredited by CreoScitex

Where Would You Use XML Schemas, As Opposed to “Plain Vanilla XML?” According to the World Wide Web

• **Publishing and syndication.** Distribution of information through publishing and syndication services. Involves collections of XML documents with complex relations among them. Structural schemas describe the properties of headlines, news stories, thumbnail images, cross-references, etc. Document views under control of different versions of a schema.

• **Electronic commerce transaction processing.** Libraries of schemas define business transactions within markets and between parties. A schema-aware processor is used to validate a business document and to provide access to its information set.

• **Supervisory control and data acquisition.** The management and use of network devices involves the exchange of data and control messages. Schemas can be used by a server to ensure out-going message validity or by the client to allow it to determine what part of a message it understands. Applications include media devices, security systems, plant automation, and process control.

• **Traditional document authoring/editing governed by schema constraints.** One important class of application uses a schema definition to guide an author in the development of documents. A simple example might be a memo, whereas a more sophisticated example is the technical service manuals for a wide-body inter-continental aircraft. The application can ensure that the author always knows whether to enter a date or a part number and might even ensure that the data entered is valid.

• **Query formulation and optimization.** A query interface inspects XML schemas to guide a user in the formulation of queries. Any given database can emit a schema of itself to inform other systems what counts as legitimate and useful queries.

• **Open and uniform transfer of data between applications, including databases.** XML has become a widely used format for encoding data (including metadata and control data) for exchange between loosely coupled applications.

• **Metadata Interchange.** There is growing interest in the interchange of metadata (especially for databases) and in the use of metadata registries to facilitate interoperability of database design, DBMS, query, user interface, data warehousing, and report generation tools.

that could be done with the information. This dissatisfaction with HTML inspired the development of the Document Object Model (DOM), which picked up where the J2008 effort left off. The DOM is a platform- and language-neutral interface that permits scripts to access and update the content, structure, and style of a document. The DOM includes a model for how a standard set of objects representing HTML and XML documents are combined, and an interface for accessing and manipulating them. With the DOM, content authors can:

1. Move one part of the document tree to another without destroying and recreating the content.
2. Create elements and attach them to any point in the document tree.
3. Organize and manipulate new or existing tree branches in a document fragment before inserting the objects back into the tree.ⁱⁱⁱ

In 1995 companies started using Internet technology to allow employees to share information and to allow customers to access their systems remotely (e.g., Intranets and Extranets). Furthermore, in 1993 and 1994, advertising was introduced to the World Wide Web and in 1995 and 1996 the first consumer commerce applications took hold. It didn't take long for companies to put “two and two together” to add business-to-business commerce applications to their Internet sites. However, HTML was inherently far too simple (read “dumb”) to deal with the complexity of commerce, with or without DOM, and SGML was far too complex for the appetites of the MIS community.

At the XML '97 conference, a simplified version of SGML was introduced: the *Extensible Markup Language*.^{iv} XML did away with the declaration by standardizing SGML notation (UNICODE), eliminating optional features such as

Omittag, standardizing the well-known angle bracket delimiters, and so forth. XML also made DTDs optional by requiring nesting of elements and employing the concept of “well-formedness”: requiring that software applications be able to deduce the structure of the document from its tagging. XML also added explicitly object-oriented features such as super typing and sub typing, and other object constructs.

XML marked the beginning of the wild frontier known as the Internet Boom (and later bust) of 1998–2000. You probably never saw XML; browsers did use XML directly until Internet Explorer version 5.0, but it was there in the background, or the server side. Where XML took root was in the interface between Web page servers (that feed your browser HTML over the Internet) and the back office systems such as company databases. The modern online catalog queries databases on behalf of customers, fetches prices and product descriptions in the form of XML data, and converts them to HTML for your viewing pleasure. If you want to make a purchase, the cached XML is used to communicate with the fulfillment and payment processing systems.

Although a lot of great Internet and E-commerce services were introduced in this period, there was also a lot of snake oil being sold to the public. Many companies and john-come-lately experts were selling the concept of E-business based upon XML, but XML was not ready nor was it appropriate for inter-enterprise computer-to-computer data exchange.

For E-commerce (read “*business-to-consumer*”) everything was fine. The XML exchange was all happening behind the scenes and was closely controlled. In some cases, a server would reach out over the Internet and pull in a banner or other relatively simple inter-enterprise exchange (e.g., Doubleclick), but any lack of rigor was expunged by that fact that there was a human on the other side of the data exchange. Even if browsers interpreted the final data somewhat differently, you (the human) could complete the interpretation.

In E-business (read “*business-to-business*”) or computer integrated manufacturing (CIM) applications of XML, there is no room for human, secondary interpretation: The whole point of E-business is to eliminate repetitive human functions (read “*re-keying data*”). XML 1.0, even with a DTD, cannot perform some basic data management functions. There is no provision for limiting field length, dealing with lookup tables of codes (e.g., product type codes), or dealing with data typing.⁴ If you have an eight-digit field for price in your purchasing database and a supplier provides you with a ten-digit price...somebody’s going to be unhappy!

So during the Internet Boom (we’ll call it “E-business round one”), a lot of E-business services and software providers claimed to use XML and they did use XML, but their XML could not necessarily be used by a system from another vendor without a lot of needless human intervention or customized programming. The snake-oil aspect was that you could buy “open software” or sign up for an “open E-business” service, but they weren’t really all that open, and many printers complained that they were being trapped into business models that they didn’t like.

XML Schema: The Door Open to E-Business Round Two

Enter XML schema. While the great Internet boom was going on, the old-time markup folks were working away at something called schema. Schema provides the parsability and data structures necessary for many aspects of E-business. (See side bar entitled, “Where would you use XML schemas, as opposed to ‘plain vanilla XML?’”) Like DTDs, as used in SGML and XML applications, schema are intended to be parsible. However, XML schema goes farther than DTDs in many aspects, including:^{vi}

- Schemas provide for primitive data typing, including byte, date, integer, sequence, SQL & Java primitive data types, etc.
- Schemas define a “type system” that is adequate for importing or exporting from data-

base systems (e.g., relational, object, and OLAP)

- Schemas distinguish requirements relating to lexical data representation vs. those governing an underlying information set
- Schemas allow for the creation of user-defined data types, which may constrain certain of its properties (e.g., range, precision, length, mask).

This last bullet is particularly important for graphic arts applications. In our industry standards and company databases we have many unique lists of codes and special data fields for everything from “product type” to “page sizes” to “color codes” and so on. Schemas allow us to define data types that are unique for our industry and let us embed requirements such as “this field shall not be greater than eight characters.” Schema also provides other features not found in “plain-vanilla” XML, such as defining substitutable elements and inheritance in parent-child data relationships. (See “Free Tutorial”)

What does this mean to you, the graphic arts professional? E-business software and systems that can read and validate schema (read “they have an embedded parser”) are in fact “open.” Once we have industry-specific standard schema for graphic arts applications, you will be better able to integrate your systems with new systems, clients’ systems, and supplier systems. But most importantly, you will finally be in a position to capture some of the ROI that was promised in E-business round one—the advent of XML schema is the opening of E-business round two, which promises less hype and more value.

Free Tutorial:

To learn more about XML schema and what they do, I recommend Roger Costello’s XML Schema Tutorial, which is available for free download from <http://www.xfront.com/>. The download (5.4 MB) includes a PowerPoint tutorial and hands-on examples and exercises. You’ll need an XML development tool that can handle schema, such as XML Spy 4.0 (mentioned above).

continued from page 5

and is in the accreditation process with other manufacturers.

KPG also introduced the Kodak Polychrome Graphics Thermal Gold Printing Plate, the next generation of the Thermal Printing Plate/830. It delivers 1–99%, 300 line screen resolution, and run lengths of one million-plus impressions when post-baked, chemical resistance, faster imaging speeds, and good ink/water balance latitude. To learn more about KPG plates visit www.kpgraphics.com.

GretagMacbeth introduced iCColor™ 200 Series, an automated spectral chart reader that sets a benchmark for productivity and price performance for use in color-managed prepress workflows. This high-precision unit can scan in both x and y directions automatically and read a chart with 299 patches in two to three minutes. Charts need only to be fed through a single time, so the operator can feed a single-sheet color chart into the instrument and literally walk away while measurements are being taken. This automatic, hands-free approach ensures precise, error-free measurements every time and eliminates the need to hand-feed multiple strips through the device.

ICColor enables prepress professionals to develop ICC profiles more quickly and with fewer errors, resulting in more consistent and reliable color matching.

GretagMacbeth also announced a partnership with **httpriint Europe** to provide underlying software and color technology. httpriint Color Management Module brings together the relevant expertise required

continued on page 8

continued from page 7

for online color management for the printing and media sectors covering the full range of color workflow requirements for remote monitor calibration, remote proofing, and other practical solutions, all supported over the Internet using http://www.gretagmacbeth.com and www.httprinteurope.com

A new Xtension for QuarkXPress called **ImageAdjuster** gives you the power to correct the color of your TIFF images inside QuarkXPress. Whether a person is an expert in color or a novice user, ImageAdjuster easily allows you to change the color of your images, giving the user greater control over color in the document. Effort and time will be saved by using the adjustment pre-sets that come with ImageAdjuster or by creating your own settings, which can be saved and reapplied to other images.

The type of corrections you can do include levels, curves, color balance, brightness/contrast, hue/saturation, selective color, channel mixer, invert, threshold, posterize, and color fill. Adjustment information created with ImageAdjuster can be opened and edited with Adobe Photoshop 6.0 or greater and supports grayscale, RGB, and CMYK TIFF images. Details for downloading can be located at www.alap.com or by contacting A Lowly Apprentice Production, Inc., Carlsbad, CA, phone is 888/818-5790.

ScenicSoft has released Color Central 4.0, their latest OPI and print management solution. This latest version supports JPEG image file formats and CopyDot files for both composite and pre-separated workflows.

continued on page 10

What Does This Mean For You?

Your direct involvement, use, and knowledge of XML will depend on what your business objectives are and how you envision your role in the printing industry of the future. Whereas most people can spend a day looking at HTML and figure it out in a nearly intuitive fashion, XML, and XML schema in particular, require training and aptitude. A few technical professionals will need to read and write XML or XML schema, more will need to be familiar with XML-enabled tools that provide embedded XML tools and more graphical and intuitive working environments, and almost all print professionals will need to understand the role of industry XML standards. Some of the ways you may choose to use XML or XML schema may include:

- Creating “house standard procedures” for selecting equipment and software so that your new equipment purchases fit an “open” architecture.
- Selecting, benchmark testing, and implementing devices that use CIP4’s JDF so that they integrate well into your plant and with your command and control systems (read “*workflow, MIS, and job management*”).
- Integrating with clients (directly or through an “E-commerce” company) using XML to avoid having to install multiple front-end systems to work with your client’s existing systems.
- Integrating your order and logistics systems with your suppliers (most likely paper) using XML to avoid having to install multiple systems to support each supplier’s existing systems.
- Providing digital asset management services to clients (internal and external) from one system, but to the multifaceted requirements of clients.
- Providing XML integration services for customers and their document creation and production systems.
- Automating the Estimate->Quote->PO/Contract->Scheduling->Tracking->Report-

>Billing cycle to minimize administrative labor and costs.

- Providing electronic document delivery and rights management.
- As an engineer or developer at a graphic arts equipment provider, you may want to learn to add XML features to your products.

Some printers will want to focus on traditional printing and will expect that their equipment and tools use XML, but other than being an informed buyer, their XML knowledge may be minimal. Others will want to obtain a competitive advantage by improving the flexibility of their production systems and providing a high degree of direct integration with key clients (read “long-term contracts and regular repeat customers”). Others will be looking exclusively for cost savings. Regardless of your major motivation, there’s an XML or XML schema either out there or in development, for you. The following list includes a few of the XML standards or XML Schema to look for. You may also want to check out the Web-XML registry of DTDs at www.xml.org/registry.

Publishing and Syndication

- **PRISM** (Publishing Requirements for Industry Standard Metadata) version 1.0: <http://www.primstandard.org/>
- **Adobe eXtensible Metadata Framework (XAP)**: <http://partners.adobe.com/asn/developer/acrosdk/docs/filefmtspecs/PDF14Delta.s.pdf>
- **ONIX**, international standard for representing and communicating book industry product information in electronic form: <http://www.editeur.org/onix.html>
- **Open Digital Rights Language**: <http://odrl.net/0.9/ODRL-EX-09.xsd>

Electronic Commerce Transaction Processing

- **PROSE-XML** (Production Order Specification/EDI XML), a transaction set for orders between printers and publishers (formerly XPP (XML for Publishers and Printers): <http://www.xmlpub.org/>

- **SPACE/XML** (Specifications for Publishers Agency Communications Exchange/XML): To be published shortly by IDEAlliance (formerly GCA), with endorsement by NAA, DDAP, and others at www.idealliance.org (Note: This may be one of the first W3C compliant schema published anywhere. Contact Mr. Craig Shrader, president of Intersect Technologies and SPACE/XML committee chair, for more information at cshrader@intersecttech.com)
- **PrintTalk**, an XML implementation for communicating quotes and orders: <http://www.printtalk.org/briefpaper04.pdf>
- **PapiNet XML Standard**, global transactions for the entire paper supply chain: <http://www.papinet.org/spec/dtd.asp>
- **PrintML** (Printing Industry Markup Language), PrintML is an application of XML that covers many messages which are transmitted between the printers, their customers, and their suppliers: http://printml.org/index_en.htm
- **AdConnexion**, Tidningsutgivarna's XML schema for newspaper advertising order exchange: write to Stig Nordqvist at sf@tu.se

Supervisory Control and Data Acquisition

- **Job Messaging Format (JMF)**, a component of CIP4's JDF (Job Documentation Format) that deals with messaging to and from graphic arts devices (i.e., presses, imagesetters, etc.): <http://www.cip4.org/>, and click on "JDF Specifications."

Traditional Document Authoring/Editing Governed by XML

- **News Industry Text Format**: <http://www.nitf.org/site/dtd.html>
- **DocBook XML DTD Version 4.1.2**: <http://www.oasis-open.org/docbook/xml/4.1.2/index.shtml>
- **Digital Talking Book** file specifications: <http://www.niso.org/Z3986.html>

- **Standard Vector Graphics**, a language for describing two-dimensional graphics in XML: <http://www.w3.org/TR/2001/REC-SVG-20010904/>

Query Formulation and Optimization

- The W3C is working on an XML Query Language (still in early development) and at present there are no graphic arts specific languages, but keep your eye on IDEAlliance's Industry Architecture Project at <http://www.gca.org/iap/>.

Open and Uniform Transfer of Data Between Applications, Including Databases

- **JDF**, the Job Definition Format: <http://www.cip4.org/>, and click on "JDF Specifications."
- **IFRATrack 3.0** is a specification for the interchange of status and management information between local and global production management systems in newspaper production: <http://www.ifra.com/>, and select "Research & Consulting" and then "IFRATrack"
- **CGATS.20/1-2001, Graphic technology: Variable printing data exchange using PPML and PDF (PPML/VDX)**: <http://www.npes.org/standards/dr10.pdf>

Metadata Interchange

- **UDEX**, IDEAlliance's Universal Data Element Cross-reference (UDEX): a component of the Industry Architecture Project: <http://www.gca.org/iap/>
- **Technical Metadata for Still Images**, image data with D.A.M. applications: <http://www.niso.org/pdfs/DataDict.pdf>
- **XMI**, Object Management Group's XML Metadata Interchange (XMI) standard: <http://cgi.omg.org/cgi-bin/doc?formal/00-06-01.pdf>

So Where Do You Start?

If you want to learn more about XML, the bookstores are flooded with XML titles. (*Buy print!*) Some recommendations:

1. A low-fluff, to-the-point **textbook** is *Learning XML*, by Erik T. Ray, O'Reilly & Associates, Sebastopol, CA, January 2001. ISBN: 0-596-00046-4 URL: www.oreilly.com/catalog/learnxml/
2. The best Internet **portal** for all things XML can be found at www.xml.org.
3. The best XML **conference and exposition** is XML 2002, scheduled for December 9-14, 2001, in Orlando, FL. See <http://www.xmlconference.net/xmlusa/> (XML, the conference and expo, is held each year in early December at different sites around the country.)
4. The best **printing-specific conferences** with XML coverage are by the Technical Association of the Graphic Arts. See www.taga.org
5. Many vendors of XML tools, from Datachannel to Arbor Text to XML Spy, offer XML-related training. For general, non-vendor specific **training**, a list of the best consultants and trainers can be found at <http://www.idealliance.org/standards/icc/>

Footnotes in order of appearance

(1) Consider what types of publications would need to transfer published information for use in future publications: technical documentation, reference publishers, and academic publishers. These types of publishers, and their service providers, have followed the development of the markup languages closely. But for the publishers and printers whose subject matter is more typically "one-off" (e.g., newspapers, magazines, POP, and general commercial printing), the markup languages may have appeared to have disappeared in to relative obscurity; and so XML to them may appear to them as more of a "new thing" than it does for the aforementioned publishers and printers.

(2) The question is often asked, is XML like C+ or Java; what can I do with it? Although XML is a "computer language" it is not a functional language that is capable of carrying out operations like the programming

continued from page 8

In addition, ScenicSoft has made it easy for third parties to add support for other file formats by introducing a new plug-in architecture. The new Color Central Sampler can create both TIFF and EPS low-resolution files for placement in your favorite desktop publishing application.

Numerous other new features and enhancements were added to this release, such as strongly enhanced support for a wide variety of different image file format compression schemes, advanced handling of clipping paths for JPEG, EPS, and TIFF files, support for QuarkXPress® colorized grayscale TIFF duotones, and the ability to maintain OPI- and font-processed files in a serve hold queue. Need more information? Check out www.scenicsoft.com.

Presstek, Inc., introduced the new Dimension auto-loading process that is specifically designed for the small- to mid-size shop that requires a CTP solution that supports multiple plate sizes and is easy to use.

Using the Presstek Automatic Plate Loading System (APL), the complete platemaking process—from imaging to cleaning to plate delivery—is accomplished with a click of the mouse. Up to twelve plates can be pre-loaded in the automatic plate-loading cassette. Plates can be of multiple sizes ranging from 9.5x12 in. to 20x20 in. for a Dimension200 to 29x29 in. for a Dimension400 equipped system. The cassette allows the user to have multiple plate sizes loaded and ready for imaging. The Presstek exclusive auto-queuing system allows the operator to specify the press on which a job will be run, and the Dimension will select the

languages. XML is a descriptive language that is used to describe data in such a way as to create information handles (or metadata) that is associated to your data and can be used to identify, validate, and manage your data openly and independent of an given application.

(3) The Department of Treasury/Internal Revenue Service's Tax Information Publications System (TIPS) was the first public implementation of ISO 8879 (SGML), which used a DTD written by Linda Brandt of the now defunct Infoconversion under contract with the U.S. Government Printing Office.

(4) To be fair, the XML specification does provide for the following data types, and XLST (XSL Transformations - 1.0, <http://www.w3.org/TR/1999/REC-xslt-19991116.html#qname>) enable users to “transform” into and out of XML, the process, without the Schema recommendation, is still not conducive to inter-enterprise exchange.

UCS-4, big-endian machine (1234 order)

UCS-4, little-endian machine (4321 order)

UCS-4, unusual octet order (2143)

UCS-4, unusual octet order (3412)

UTF-16, big-endian

UTF-16, little-endian

UTF-8

Bibliography

ⁱ Cover, Robin, *The XML Cover Pages*, SGML/XML Bibliography Part 4, I-L, OASIS, <http://www.oasis-open.org/cover/bib-il.html#iso10744>, August 3, 1999.

ⁱⁱ Harvey, James E., *SGML Applied to Automotive Service Information*, CALS Journal, Saratoga, NY, Fall 1993, pp. 27–31.

ⁱⁱⁱ *Document Object Model Overview*. MSDN Online Workshop, Microsoft Corporation: <http://msdn.microsoft.com/workshop/author/dom/domoverview.asp>, June 2000.

^{iv} *Extensible Markup Language (XML) 1.0*. W3C Recommendation Feb. 10, 1998. <http://www.w3.org/TR/1998/REC-xml-19980210>.

^v *XML Schema Requirements*. W3C Note February 15, 1999, <http://www.w3.org/TR/NOTE-xml-schema-req>.

^{vi} *XML Schema*. W3C Recommendation, May 2, 2001, Parts 0-2, <http://www.w3.org/TR/xmlschema-0/> (Primer); <http://www.w3.org/TR/xmlschema-1/> (Structures); and <http://www.w3.org/TR/xmlschema-2/> (Datatypes).

James E. Harvey is the founder and President of Media4theWorld, LLC, which provides business development services and management and technical consulting to various software companies, the Technical Association of the Graphic Arts (TAGA), and IDEAlliance. Previously he was vice president of Spectrum Operations at Graphic Communications Association (now IDEAlliance) where he was the founder of the Industry Architecture Project and GRACoL, and was responsible for SPACE, managed the CtP Users Group, and coordinated Spectrum conferences, among other responsibilities. He has held many industry posts including VP & Board member of TAGA, secretary and member of the executive committee of the Digital Ad Lab, and secretary of ANSI/PIMI IT2-28 (Densitometry). Jim can be contacted via email at jharvey@media4theworld.com or via phone at (410) 451-5040.