Paperless Technologies to Save our Planet... one Tree at a time.

IFD Management Report

Advantages and Disadvantages that You Should Know About XML

Author

Dan Forster
President
inFORM Decisions, Inc.

inFORM Decisions
Leaders in Document Automation

inFORM Decisions, Inc. • 30221 Aventura • RSM, California 92688
949.709.5838 • www.informdecisions.com

IBM Business Partner

Supporters of the Arbor Day Foundation™

XML

Extensible Markup Language (**XML**) is named such because users of it can define their own markup elements. It can be used to encode documents and to serialize data. Its primary purpose is to facilitate the sharing of structured data across different information systems. XML is becoming the standard for the electronic exchange of data between businesses and on the Internet.

Advantages

There are distinct reasons to consider adopting XML at your installation.

Advantage 1 – Readability: It is text-based and therefore human readable. Moreover, given the use of text tags to demarcate the data, data represented by XML is usually somewhat understandable to the reader without reference to any other file or definition.

Example: to send a simple database file containing a user account number and balance, the XML could look something like this:

```
<file>
  <customer>
    <name>John Smith</name>
            <name>Gail Smith</name>
    <acctno>0100001234</acctno>
    <currentBalance>1.00</currentBalance>
  </customer>
</file>
```

Essentially the 'customer' tag wraps what would be a row in a database. Each of the database fields are elements in the XML file. There are two names associated with this account, not an uncommon occurrence, although in a database file the two name fields would have to have different field names, e.g. name1 and name2.

Data that is specified as an element enclosed by tags and that is unique to its parent element can be specified as an attribute of the parent element. So one could have this:

```
<file>
  <customer acctno=”0100001234” currentBalance=”1.00”>
    <name>John Smith</name>
    <name>Gail Smith</name>
  </customer>
</file>
```

Clearly most IT professionals would have little trouble immediately comprehending the contents of this XML file. No external definitions need to be referenced.

Advantage 2 – Universality: It supports the Unicode Standard, so text from any character set can be used. Moreover, the text and elements can be specified in different character sets. For example, the element wrappers could be in Sanskrit with the data in these elements in Japanese. Or the element wrappers could be in our Roman character set and the data in the elements in Russian Cyrillic.

Advantage 3 – Binary Representation: XML can store binary data. The data must be converted to text. This is usually done with algorithms such as Base64. This is the same algorithm used to convert binary data to text for the MIME standard for email. Wikipedia gives an example of the calculations involved in transforming the letters 'man' to 'TWFu' using Base64. The result is that three characters of input data are converted to 4 characters of output.

Also, binary characters can also be separately represented by numeric character references. These have the format:
&#x
followed by the 4-character alphanumeric hexadecimal code for the data. Four characters are required to represent Unicode. A semi-colon terminates the character. Example:
&#x20AC;
represents the symbol for the Euro monetary unit.

Commonly used characters can also be separately represented by entity references, which begin with an ampersand (&) and follow with the entity's name terminated by a semi-colon. So for example, the '>' character is represented as an entity reference by:
&gt;

Advantage 4 – Data Storage: XML will store any document or data that can be represented as a tree structure. DB2/400 databases are more easily represented as grids. Each row has the same structure and within each row each field name is unique. XML by virtue of its structure permits multiple instances of elements within a parent element, which is equivalent to having multiple instances of a particular field within a database record. In the example above, the account record has two name elements, which if it were a DB2/400 database would mean having two identically named fields in the same record. See last disadvantage below for limitations.

Advantage 5 – Browser Display: XML can be displayed on browsers. Use of an XML Stylesheet can control this display. Without the stylesheet, which is a separate file, a browser will usually display an XML file as a collapsible list / tree. If you want to see such a file in practice, open the RSS feed from a web site in your browser. The RSS file is in XML, and the browser will probably render it as a collapsible list.

Advantage 6 – Platform Agnostic: A major benefit of XML is that it is platform independent. XML is basically text. When moved between platforms, the characters set translation is transparent to the user. This is similar to the ASCII FTP mode with which most are familiar. XML is recommended by the World Wide Web Consortium. It is a fee-free open standard.

Advantage 7 – SEC Compliance: An important variant of XML is XBRL (Extensible Business Reporting Language) which is recommended by the United States Securities and Exchange Commission (SEC) for reporting financial earnings information to the U.S. Federal Government. Chris Smith, Senior Editor at MC PressOnline reports, “the SEC, which for the past two years has been accepting voluntary XBRL financial filings, now has a roadmap leading toward mandatory filing by this Fall (2008) for Fortune-500 class corporations. For more on XBRL, and other uses of XML, see References below.

Disadvantages

Some cautionary observations are important to recognize before you adopt XML.

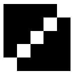
Disadvantage 1 – Inefficient: XML was not originally defined as a database storage platform. It was designed to accommodate the exchange of data between nodes of dissimilar systems. Compared to other storage algorithms XML is relatively inefficient. The XML tags, which make it readable to humans, require additional storage and bandwidth.

Example:
Binary formatted file account number: 00100001234, or 11 bytes.
XML formatted file account number: <acctno>00100001234</acctno>, or 28 bytes.

Disadvantage 2 – Inefficient Binary Storage: Storage of binary data, such as image data, is also inefficient. Storage is frequently done using Base64 encoding (also used for MIME attachments in email). This requires four text bytes to store three bytes of binary data, taking up more storage and bandwidth. Representations using named character references or entity references of one or two byte characters can require as much as eight bytes.

Disadvantage 3 – Encoded Image: An additional program is required to display encoded image data represented in XML. The encoded binary image data appears as a stream of unintelligible character data. To display it, it must be un-encoded and then reassembled into an image.

Example: Consider the basic raster 5 x 5 image below.



This is a black/white image that looks like an enlarged forward slash. When stored using numeric character references, it could look something like this in the XML file:

&#00;&#00;&#00;&#00;&#01;
&#00;&#00;&#00;&#01;&#00;
&#00;&#00;&#01;&#00;&#00;
&#00;&#01;&#00;&#00;&#00;
&#01;&#00;&#00;&#00;&#00;

Converted to a numeric grid, it could be represented thus:

0 0 0 0 1
0 0 0 1 0
0 0 1 0 0
0 1 0 0 0
1 0 0 0 0

But to display it as it was shown in the basic raster 5 x 5 image above, a black square containing white squares within, requires an image-rendering program.

Disadvantage 4 – XML Parser: There are three general methods to parse XML files. Most programmers without any experience in XML will be unfamiliar with all three:

> **Programs**: Write a program and invoke the SAX APIs. These APIs are callbacks that are triggered at certain points in the parsing. The developer specifies the methods invoked for these callbacks.
>
> **APIs**: Use the DOM API, which treats the entire XML document as a tree of data nodes. The program can traverse the tree processing the data.
>
> **Engines**: Use a transformation engine and filter to transform the document for display/print.

Disadvantage 5 – XML Stylesheet: XML can be used to transmit and store documents for visual display. However, storage is not its primary or original purpose. To use it for anything beyond basic display of the data using markup requires an additional file – the XML Stylesheet. XML Stylesheets are conceptually identical to HTML CSS files, but with a different syntax. For more: http://articles.techrepublic.com.com/5100-22-5435275.html

Disadvantage 6 – Backend Technology: XML is ultimately a backend technology in the form of data storage and transmission technology. It lacks rendering instructions. Documents destined for display to the end user are better held in Adobe's Portable Document Format (PDF) or other similar technology where the rendering is completely described. PDF file format is in part a subset of the Postscript page description language.

Note: The Adobe Mars Project aims at an XML-friendly representation for PDF documents. This XML representation of PDF documents makes the storage, construction and retrieval of parts of a PDF document easier. The project uses XML to store PDF data. More information.

Disadvantage 7 – Tree Structure Problems: Not all files are well-represented with tree structures. Most DB2/400 tables are well-represented as grids. Moreover, flat files and *undefined fixed record length files* are usually portrayed as grids. Finally, *stream files* from the IFS are unstructured unless the developer has written structured data into them. Stream files are unlikely to be meaningfully represented using the tree structure required by XML.

Summary

XML provides a way to transfer and store between different nodes in a network or across the Internet. It is easy to learn and use as well as intuitive to view. The technology is blessed with a mixture of strengths and weaknesses. In assessing whether to adopt this technology, it is important to assess how both the strengths and the weaknesses relate to your own installation. inFORM Decisions has technicians that can help you with this process.

It’s important to remember that for files where the intention is visual display, other file formats such as Adobe PDF are usually a better choice.

XML is sometimes mandated for use. In this case, the material contained in this IFD paper should be of assistance in understanding how to get the best out of the XML technology you are about to implement.

Finally, a variant of XML, Extensible Business Reporting Language (XBRL) is becoming the de facto international standard for complying with mandated financial reporting by organizations including the Securities and Exchange Commission.

References

Chris Smith, Senior Editor, *A Decade in the Making, XBRL is Poised for Adoption*, Jan. 28, 2008. http://www.mcpressonline.com/web-languages/a-decade-in-the-making-xbrl-is-poised-for-adoption.html

Jim Barnes, *RPG IV and XML Together*, Nov. 1, 2000.
http://www.mcpressonline.com/rpg/rpg-iv-and-xml-together.html

Jim Barnes, *Fast XML with RPG IV and SAX*, May 1, 2001.
http://www.mcpressonline.com/rpg/fast-xml-with-rpg-iv-and-sax.html

Chris Green, *XML Part 3: Code Development*, Feb. 9, 2004.
http://www.mcpressonline.com/web-languages/xml-part-3-code-development.html

Mike Faust, *Access XML Services with JaveScript*, Oct. 26, 2007.
http://www.mcpressonline.com/scripting/techtip-access-xml-web-services-with-javascript.html

Marie Wilson, *Getting Started with DB2 Extender for iSeries*, July 21, 2003.
http://www.mcpressonline.com/web-languages/getting-started-with-db2-xml-extender-for-iseries.html

Jarek Miszczyk, *A Fast Path to XML Integration with DB2 for iSeries*, Nov. 28, 2005.
http://www.mcpressonline.com/web-languages/a-fast-path-to-xml-integration-with-db2-for-iseries.html

Extensive XML tutorial.
http://www.w3schools.com/xml/default.asp

A Technical Introduction to XML
This introduction to XML is geared to a reader with some HTML or SGML experience, although that experience is not absolutely necessary.
http://www.xml.com/pub/a/98/10/guide0.html

About inFORM Decisions

Since 1994, inFORM Decisions has developed secure document automation solutions for IBM midrange users. An Award Winning IBM Business Partner for over 10 years, the company pioneered the concept of eDocument output routing (SmartRouting^{TM}), with user-defined business rules automating workflow in the enterprise.

IFD's SmartRouting uses user-defined conditions to route designed forms, reports and checks through email, fax, or conditional printer OutQues, or simply archives these eDocuments in a Web-based portal. Enterprise Content Management solutions from IFD represent end-to-end document automation that runs native on IBM AS/400, iSeries, i5 and System i servers. The company's Document Automation family includes: **iDocs**, **iMail**, **iFax**, **iPDF** and **iView**. Banking Automation solutions include **iChecks**, **Check-Secure**, **ACH-ePayment** and **Positive Pay** laser check printing.

Demo versions of these software solutions can be downloaded at:
www.informdecisions.com/download

The company is a supporter of The Arbor Day Foundation and other forest protection agencies and continually expands its efforts to have a positive impact on the environment.

Corporate headquarters is located in the city of Rancho Santa Margarita, Orange County, California. For more information: www.informdecisions.com.

Other Materials from IFD

IFD continually promotes the general awareness and benefits of a paperless operation, both for the organization and the environment.

The company's own green philosophy is described in its **IFD Green Page** along with an extensive list of Web resources related to the greening of the office and the data center: www.informdecisions.com/trees.

The company contributes journal articles to foster a greater understanding of the value of document automation. Some recent examples:

> The Green Revolution Hits the Office, DOCUMENT magazine's *eNotes* newsletter, December 17, 2007.
>
> The Automated Clearing House (ACH) Network -- A Panacea for Electronic Payments and eBanking, MC PressOnline, July 26, 2007.

A number of white papers authored by the company are available from the IFD download page at: www.informdecisions.com/download. Papers on document automation and MICR check printing are available, plus a self-assessment on document automation readiness.

Members of the organization are available for public speaking on topics related to the paperless revolution.

inFORM Decisions
Leaders in Document Automation