

Using extensible mark up language for the single source delivery of teaching resources via print and the Web: A practical example

Lucas Walsh

*School of Communications and Multimedia
Edith Cowan University*

This paper seeks to provide an introduction to Extensible Mark up Language (XML) by looking at its use in a single source publishing approach to the provision of teaching resources in both hardcopy and online. Using the development of the International Baccalaureate's online Economics Subject Guide as a practical example, this discussion reflects on some of the key technological and educational limitations and strengths of using XML.

Keywords: XML, publishing, International Baccalaureate Organisation, IBO, Online Curriculum Centre

Introduction

Between 2001 and 2003, the International Baccalaureate Organisation's¹ Online Curriculum Centre developed and implemented Extensible Mark up Language (XML) to deliver teaching resources in both printed and Web based modes of delivery. XML was introduced so that a single source of content could be used in different ways and formats according to the needs of both content providers and end users.

The pilot project to test the technical and educational effectiveness of this process of "single source publishing" was the Economics Subject Guide for the IB's Diploma Programme.² Single source publishing is a process whereby a master source of information is prepared in an authoring and editorial process, and then used to drive a production process to delivery to end users through mass print and the Web. Using this approach, paper based documents would look and be structured as normal; however, the online version of the same content would be reformatted in an automated process to be more interactive and Web friendly.

The use of XML within this kind of database driven framework has developed tremendously during the last few years; however, many educators are still coming to terms with how to best apply this technology to a "learning object approach" to the delivery of educational resources, so that learning content is reusable, interactive and can be customised by the end user.³ The following discussion explores some of the issues arising from the use of single source publishing to drive the delivery of printed and online curriculum resources to IBO teachers. Using the development of the Economics Subject Guide as a practical example, this paper reflects on some of the key technological and educational limitations and advantages of using XML.

The first section of the paper sets out the context of discussion, in which a single source publishing approach was adopted by the *Online Curriculum Centre* (OCC) to enhance the way teaching resources were delivered to teachers of the International Baccalaureate (IB) throughout the world. Part 2 provides an introductory overview of XML. Part 3 describes the workflow process required to repurpose the Guide using this approach. The fourth section looks at the challenges and limitations encountered when implementing XML. The final part of this paper looks at the technological and educational strengths of this XML based single source approach.

¹ Founded in 1968, the International Baccalaureate Organization (IBO) is a non-profit educational foundation that delivers an internationalised curriculum to children and young people aged 3-19 across three programmes: the Diploma Programme, the Middle Years Programme and the Primary Years Programme. As of July 2004, 1,381 authorized IB schools offered a total of 1,634 IB programmes in 117 countries to approximately 200,000 students (IBO 2004).

² The IB Diploma Programme is a pre-university course of study that leads to examinations designed for secondary school students aged 16 to 19. The Programme, which incorporates elements of national systems, is a two-year international curriculum available in English, French and Spanish, that generally allows students to fulfil the requirements of their national or state education systems.

³ The term "content" is used here to include all text, illustrative and audio-visual materials delivered via printed hardcopy and the Web, such as Subject Guides, teaching support materials, and general educational resources.

While the purpose of this paper is to serve as an introduction for educators considering the use of this kind of database driven approach to e-learning and the provision of educational resources, a basic understanding of HTML is assumed by the author. But I also ask the reader to bear in mind that my own understanding of the technology is first and foremost from the perspective of an educator, rather than a programmer or Web developer. Nevertheless, the reader is asked to be patient with the frequent use of acronyms, which I have summarised in the table below:

Table of Acronyms	
CSS	Cascading Style Sheet
DTD	Document Type Definition
FOP	Formatting Objects Processor
HTML	Hypertext Mark up Language
IB	International Baccalaureate
IBO	International Baccalaureate Organisation
OCC	Online Curriculum Centre (IBO professional development website)
PDF	Portable Document Format
SGML	Standard Generalised Mark up Language
W3C	World Wide Web Consortium
XML	Extensible Mark up Language
XSL-FO	XSL Formatting Objects
XSLT	Extensible Style sheet Language Transformation

Context

When the Online Curriculum Centre commenced development of a single source publishing approach using XML, its website was seeking to deliver education resources and services in English, French and Spanish to over 25, 000 IB teachers working in around 1,300 authorised IB schools across more than 100 countries.⁴ Originally established in 1999, this password protected website was developed in response to the expansion of the IBO across the world and the need to disseminate information quickly and efficiently. It was designed to support the IBO curriculum in all three academic programmes, incorporating a database of educational resources to utilise the main strength of the Internet by providing IBO teachers with the facility to communicate and share their knowledge within and across schools and national borders. The site enabled teachers to: view curriculum materials; exchange details of useful resources (e.g. websites, books, magazines) and teaching ideas; discuss curriculum issues with other teachers from around the world; take part in online training through bespoke, curriculum specific educational modules; and keep up to date with IB curriculum developments and general news about the IBO.

Redevelopment of the new OCC website began in the latter half of 2001, which involved overhauling its frameworks for the delivery of Subject Guides, teaching materials and a-synchronous discussion forums. Redevelopment of the site attempted to address the delivery of an estimated 3.2 million words of content (e.g. Subject Guides, general news and information) targeting three areas: (i) the storage of data - in particular, how curriculum documents and other teaching resources were stored; (ii) the dissemination of data - how documents were prepared and delivered to users (IB teachers, school administrators and support staff) via both print and the Web; and (iii) the presentation of data - how the documents looked both on paper and on the Web, which at the time was generally the same because the majority of content on the OCC website was taken directly from printed source files stored in PDF.

A major component of the renovation of the OCC was to develop a framework for publishing online versions of curriculum materials, redesigning them from their original format for paper publication. The goal was to develop a framework in which content could be repurposed (i.e. restructured and reformatted) from standard printed publications to the Web in such a way that maximises the full power of the Internet medium. The storage of OCC data, such as curriculum documents and other teaching materials on the IBO's internal computer network, came from multiple sources and formats (e.g. PDF, HTML, MS Word and WordPro), which lead to inefficiencies in data management and unnecessary duplication of labour. Previously, teaching resources such as Subject Guides were prepared in printed format (typically standard

⁴ Such was the rapid expansion of the IBO that the numbers of IB teachers and authorised schools steadily increased during the 17 month period covered in this paper.

black and white A4 sized documents), then marked up in HTML in a separate process for the OCC website. The layout and format of online versions were predominantly text based and looked similar to the printed versions. Often, the Portable Document Format (PDF) used for printing was directly uploaded to the Web as a single file. In some cases, these PDF files designed for printing purposes were often scanned off a master document and posted directly to the website in a format that was sometimes difficult to view via a Web browser - especially when these PDF files were extremely large in size (particularly those containing images). Furthermore, these files were typically 'static', in the sense that they were not designed for interactivity (e.g. many did not include hyperlinks or navigation facilities). This static way of storing content in PDF and HTML was not dynamic and underutilised the power of the database driving the OCC website. Furthermore, raw content was formatted using different word processing, Web design and other packages without a single standard approach.

A system was needed that 'captured' content early on at the authoring and formatting stage so that the content could be marked up for print and the Web in a unified process. Consequently, a single source publishing approach using XML was adopted. The primary aim was to enable teachers to be able to access two differing versions of syllabus, each with separate function. The Diploma Programme Subject Guide for Economics was used as the proof of concept for this approach. This Subject Guide consisted of around 7,000 words of text outlining the objectives, syllabus and assessment requirements for the subject. The IB Economics curriculum was taught at two levels, Higher Level (HL) and Standard Level (SL), as outlined in the Subject Guide (IBO 2003). Though published in English, Spanish and French, this paper will focus on the development of the English language version of the Economics Guide.

A brief introduction to XML

XML⁵ is a set of rules and guidelines for designing text formats that enabled content providers at the IBO to structure, store and deliver data (such as Subject Guides) to teachers via the Internet and in printed hardcopy formats. XML is a text based mark up language that is 'extensible', or customisable, to create other languages (Bradbury 2001). It is not a mark up language in itself, but rather a system for defining specialised mark up languages used to transmit formatted data. This "metalanguage" (i.e. language that can create other languages) was developed in response to some of the positive and negative experiences of other languages, such as Hypertext Mark up Language (HTML) and Standard Generalised Mark up Language (SGML). Developed so that contractors and suppliers could exchange big technical documents in areas such as the US Department of Defence, pharmaceutical and aerospace industries, the first working draft of SGML was published in 1980 (Kasdorf 2003, p. 74-75). Development of XML began in 1996 under the auspices of the main standards body for the Internet, the World Wide Web Consortium (W3C), as a subset of SGML that would be easier to implement on the Web (Kasdorf 2003, pp. 75-76). XML's designers drew from: (i) the positive aspects of SGML; while (ii) seeking to address some of the negative aspects of HTML, which had a lot of limitations as a mark up language (Kasdorf 2003, p. 90). XML arose from the need to develop something simpler and clearer to use. By 1998, W3C recommended XML as a standardised approach to data handling (W3C 2003).

Kasdorf writes that XML "provides a well defined, broadly accepted syntax for creating markup schemes to [enable] the development of tag sets focused on projects as small as a single book or journal... or whole collections or classes of books or journals... or to achieve a particular functionality" (Kasdorf 2003, p. 87). It consists of Elements (mark up that encloses content), attributes (which embellish content), comments (such as human readable annotations) and entities (which enable any chunk of digital content to be given a name and then referred to by that name instead of being directly incorporated into the XML document itself. These chunks can include content of external files, including non-XML files (Kasdorf 2003, p. 101).)

Rather than replace HTML, which is primarily a presentational format "designed for a specific way of viewing the given content" on the Web, XML is a "language designed for marking up content of all sorts (text, graphics, computer code, commercial information, intellectual information) in order to specify what the component parts of that content are, and to describe them" (Kasdorf 2003, p. 91). Like HTML, XML makes use of tags (words bracketed by '<' and '>') and attributes (of the form name="value"). Unlike HTML, XML has a potentially infinite number of tags. HTML, for example, is one of the many tag sets that can be defined in XML. Data stored in an XML file usually consists of strings of words, rather than

⁵ The acronym XML not only refers to the language itself, but also to a whole family of standards that have developed around it, such as XSL, which is outlined below (Kasdorf 2003, p. 88).

numbers. These words can have explicit meaning assigned by the creators of the content. As Kasdorf (2003, p. 91) explains, “XML is, in its best use, a source markup, designed to be transformed into something else, to be archived and used in different ways. In fact, HTML is a frequent output from an XML archive. XML retains the information about what the elements in the source content are and how they relate; HTML typically loses that information, focussing instead on what the content should look like in a given presentation.”

Where HTML specifies what each tag and attribute means and often how the text between these tags will look in a browser, XML uses the tags only to delimit pieces of data, and leaves the interpretation of the data completely to the application that reads it (Bradbury 2001). For example: use of the “<p>” tag in an XML file does not necessarily define a paragraph as it does in HTML. In XML, it can signify any attribute. In relation to the Economics Subject Guide, these tags were created to suit the specific needs of the content providers and in a language that participants in the creation and dissemination of the Guide could understand (i.e. technical developers, designers, educational authors and lay persons).⁶ Any tag can be created and used in XML, provided it is used in a pair with a closing tag to match any opening tags – unlike HTML, which will typically attempt to display a webpage even though it may be missing certain tags, XML will not tolerate any mistakes; consequently, browsers may display inappropriately tagged data as a ‘jumble’ of data. This “draconian error handling” is deliberately designed to promote stability in data handling (Kasdorf 2003, p. 91).

Workflow

For the Economics Subject Guide, XML was not only used to generate the HTML version for the Web, but also to generate fully formatted PDF versions of selected portions of the same source of content, which would be mass printed and made available to teachers in hardcopy. The basic steps for repurposing the Economics Guide for viewing on the Web and mass printing were as follows:

Step 1: The Economics Guide was created in Microsoft Word, through which the basic styles (e.g. for topic headings) were formatted. Word was used because the software was familiar to authors;

Step 2: The document was then structured and tagged according to a Document Type Definition (DTD). A DTD is a kind of template for the XML; it is a separate file that defines what should go where within an XML document and in what order. A number of DTDs have been defined and released to the community to use. For the Economics Subject Guide, a specific language was created and defined as a DTD based on a standard *DocBook* Document Type Definition. *DocBook* is general purpose XML DTD specifically suited to book and journal publishing. It features a simple schema that served as a good starting point because it was available freely and applicable to the Subject Guide, in that it was not very complex in structure and had a relatively simple format. This DTD stated what elements must be present, which ones were optional, what their attributes were and how they were structured in relation to each other. The XML document containing the Economics Guide was written to comply with this DTD to ensure that the document could be processed for display on the Web. YAWC software⁷ was used to insert DocBook tags into the Microsoft Word document to define the structure of the Guide (e.g. sections, heading levels, tables, lists, etc.);

Step 3: The document was then validated to ensure that it conformed to the DTD structure;

Step 4: The next step involved exporting / converting the Guide to the “raw”, single source XML document that served as the source for both online and printed versions. Having styled the content in Word format using *DocBook* tags, the Guide was then exported from YAWC to create a basic, structured XML document. (Essentially, Word styles were replaced with XML tags.) This XML was then refined using XMetaL software to define in more detail certain features of the XML document, such as roles, attributes and table widths for the layout of the Guide.

This XML document also served as a container for metadata; that is, information about the content in the Guide. Metadata was added consisting of attributes and values that identified, for example, the language

⁶ For example, the tag for content relevant to Higher Level (HL) teachers of IB Economics, was: section role=”hl”. For sections of the Guide relevant to both Higher Level and Standard Level (SL) teachers, the tag defined this portion of content as: section role=”hl,sl”. These tags enabled manipulation of how Higher and Standard Level content appeared online. A Higher Level section of the Guide could, for example, appear as highlighted text for users logged into the website as Higher Level Economics teachers.

⁷ YAWC (Yet Another Word Converter) is a plug-in for Microsoft Word that enables Word documents to be converted into HTML or XML.

in which the Guide was written (i.e. English, Spanish or French), the author, the teaching level (HL or SL), publication date, security, rights information, version and so forth. Other metadata was added describing how the content of the Guide would appear and relate to other online content (such as the online discussion forums). For the Online version, this metadata enabled the Guide to be customised to deliver the relevant material to end users, pitched at a teaching level based on a user profile stored in the OCC system. For example, when a Higher Level Economics teacher requested the Subject Guide online, the database driving the website was programmed to find the data tagged at that teaching level and covering only the subject matter required in the teacher's working language (English, Spanish or French). The system was programmed to find the data tagged at their teaching level, visually highlighting the teaching level of subject matter specified. If, on the other hand, a print out was requested, the entire Guide, including both Higher and Standard Levels, was automatically generated from the same single source.

Metadata was further added to the online version of the Guide so that new features, such as graphics, illustrations (e.g. Demand and Supply Curves) and other features could be hyperlinked to relevant sections. Keywords and headings were identified and linked to a definition within a glossary. Selected aspects of the syllabus could, in future, be extracted from the single source in a summary form for the teachers to distribute to students. The Subject Guide also incorporated the use of hypertext to link to a facility for adding resources. Furthermore, key words, phrases and headings were linked to other resources which could be added by users to the relevant area of the Subject Guide. Users were able to add different types of resource to selected keywords and headings. Resources included information about books, websites and teaching ideas, such as course outlines, lesson plans, student activities, tests and quizzes. Resources such as specimen questions, further diagrams, glossary and command terms were added later to the online only version of the Guide by clicking on specific words or phrases within the syllabus. The Subject Guide was also tagged to enable portals to other related resources on the World Wide Web (e.g. Biz/ed 2004). A holistic approach was taken in an attempt to integrate the Guide with the OCC website's other facilities, such as the discussion forums.

Step 5A: The XML file containing the Subject Guide was then converted into HTML for viewing on the Web using an Extensible Style sheet Language Transformation (XSLT). XSLT is an XML based language for transforming XML into other formats. (Bradbury 2001). For the Guide, XSLT was used to take the XML file containing the Subject Guide and convert it to HTML. The XSLT style sheet rules ensured that certain "online only" features were recognised in the XML document (such as the Subject Guide's glossary of economic terms, graphs and other illustrations):

Step 5B: While one XSLT was used to create the online version, another was used to create a printed version (in conjunction with XSL-FO to generate the PDF) using the same source file in XML. This XSLT style sheet effectively converted the same XML file to a format for print (FOP⁸). The style sheet ensured that online only tags were ignored for the printed version. A Java application, XEP, was used to render the content in PDF. The Guide was then printed for mass distribution to IB schools.

XML was used within this single source approach for four main reasons: (i) to encode and structure information from one source; (ii) to enable the format, or 'look', of the Guide to vary according to where and to whom it was being delivered (be it in printed hardcopy or via the Web); (iii) to exploit the openness of XML standards; and (iv) though not fully exploited during the pilot phase, XML can be used as a storage medium. To understand how XML works, it is worth explaining these reasons in further detail.

XML as an encoding mechanism

Firstly, XML was used as an encoding mechanism to structure and manipulate information. It was not designed as a programming language; rather, it was developed to improve the way computers generate data, read data and ensure that the data structure is unambiguous and more stable. So, XML provided a set of rules, guidelines and conventions for designing text formats that enabled the structuring of data.

⁸ FOP stands for Formatting Objects Processor. See Apache's (2004) website for more information. This process for Extensible Style-sheet Language Formatting Objects (XSL-FO) provided a way of creating paginated print versions of data stored in XML documents. XSL-FO is a page description language which tells a formatter how to structure text on a page (e.g. specifying size of margins). The source XML document was transformed into FO using a style-sheet and then rendered as a file format suitable for printing, such as PostScript and PDF (Kasdorf 2003).

Using XML to divorce display from content

XML was useful for storing information that needed to be presented in a variety of formats. Using XML, it was possible to divorce display from content using separate style sheets. As mentioned above, the same source of content used for the printing process was used for the Web; however, both outputs used their own XSLT so that they looked and functioned differently. Figure 1 offers a visual comparison between the printed and online versions of the same source of content.

Printed version	Online version
<p style="text-align: center;">Section 3: Macroeconomics</p> <p>The purpose of this section is to provide students with the opportunity to explore the major macroeconomic issues facing countries' economic growth, economic development, inflation and income distribution.</p> <p>Section 4 deals with external equilibrium. Income distribution is addressed in greater detail in section 5.</p> <p>The economic strategies available to governments—demand-side intervention—are introduced and evaluated. These policies are macroeconomics, international economics and development economics.</p> <p style="text-align: center;">3.1 Measuring national income</p> <ul style="list-style-type: none"> • Circular flow of income <p style="text-align: center;">(Source: IBO 2003a)</p>	<div style="border: 1px solid black; padding: 2px; margin-bottom: 10px;">Section 3: Macroeconomics</div> <p style="text-align: right; color: green;">Syllabus details</p> <ul style="list-style-type: none"> ▶ 3.1 Measuring national income ▶ 3.2 Introduction to development ▶ 3.3 Macroeconomic models ▶ 3.4 Demand-side and supply-side policies ▶ 3.5 Unemployment and inflation ▶ 3.6 Distribution of income <p style="text-align: center;">Section 3: Macroeconomics</p> <p>The purpose of this section is to provide students with the opportunity to explore the major macroeconomic issues facing countries' economic growth, economic development, inflation and income distribution.</p> <p>Section 4 deals with external equilibrium. Income distribution is addressed in greater detail in section 5.</p> <p style="text-align: center;">(Source: IBO 2003b)</p>

Figure 1: Divorcing display from content: Two outputs from the same source

Cascading Style Sheets (CSS) were used with the Subject Guide's XML file to turn its XML tags into layout instructions (specifying the fonts, backgrounds, colours, etc.). An instruction was written into the XML file specifying how it should use the CSS file to define the way Web browsers displayed the file online. Changing the whole look of the Subject Guide only required a change to the CSS file, rather than to the whole document. If another website was going to display the Subject Guide, they could, in theory, apply their own CSS file to the file to make it look more consistent with their own website (the IBO uses a number of websites for different purposes and audiences). Using XML, it became possible to also specify that only parts of the Guide appear on certain websites. Selective versions of the Guide could be created dynamically on demand in response to a user's / website's request. The content could then appear differently on various IBO and external websites, customised to the needs of their respective users and according to the design motif adopted by the respective websites.

If required, XML could in future also be used for multi-channel display systems because it can be converted to a variety of different display formats, such as interactive television, PDAs and mobile phones. HTML versions of content can be made more accessible to users with physical impairments and other special learning needs. The same source of content could be made available to speech based applications, Braille, Wireless Mark up Language (WML) for mobile devices, by writing another XSLT transformation to convert the source to the required output. XML can, in principle, be used to serve different clients using a 'one size fits all' system (Bradbury 2001).

The capabilities of using XML to drive a single source publishing process are summarised below in Figure 2, which has been adapted from a digital publishing schematic referred to as "centralised content aggregation" (Merceica 2001, p.81). This illustration shows how a single source of content can be used across multiple outputs using XML:

XML as an open standard

A major advantage of using XML is its openness. XSLT and XSL-FO are based on open standards, which were adopted in part because the software was freely available. XML can actually be authored in any basic word processing package, reducing the need to depend on specific vendors or software. Because XML has become a pervasive standard for open information interchange, it meant that the organisation did not have to be overly dependent on commercially developed proprietary, closed standards that can be

extremely expensive. Nevertheless, it was still necessary to use PDF, which is a widely adopted page description language but remains a de facto standard controlled by Adobe Systems (Kasdorf 2003, p. 76).

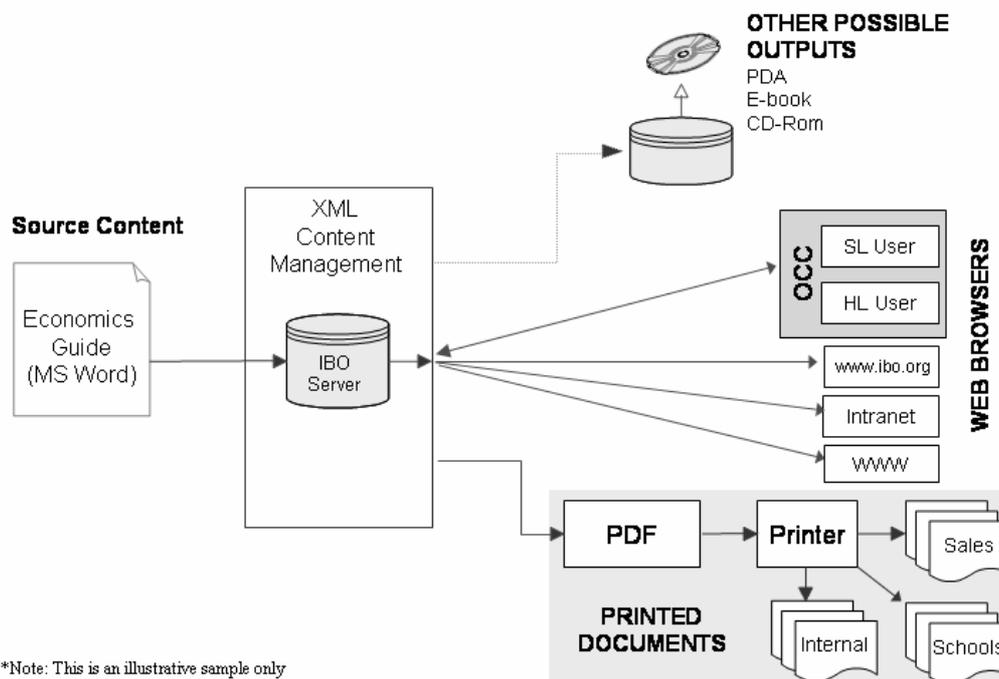


Figure 2: An example of single source publishing

XML as a storage medium

Although the Guide was originally mastered in MS Word, the source XML ‘version’ could then be used as the data storage medium from then on. A challenge for many organisations collecting large amounts of information is maintaining content. XML is useful for content management, because it can be used to tag data for customised storage (Bradbury 2001). It is possible to store all sorts of information about a data item or learning object within its XML file.

Limitations of XML

Adopting XML involved an extensive process of implementation. As shown from the basic steps described in Part 3 above, the process of preparing content in XML was quite complex. The initial outlay of human resources and finances for this project was considerable, not only because of the new technical requirements for XML, but also because development of the online interactive Economics Subject Guide represented a radical departure from conventional processes and workflows. Extensive in-house and external training was required in handling content mark up in XML. The OCC Web developer and staff responsible for print based publication needed to undertake training in XML and XSLT, as well as in new software packages, such as YAWC and XMetaL. An external consultant was also required to assist development and implementation of XML. Familiarisation with the basic standards of XML was required by all of those involved in the process.

A lot of work was required to develop XML elements within the scheme of data management and formatting of the Subject Guide. Deciding on which content was generic required careful consideration (i.e. which content would be common the all output versions of the Guide, as well as potentially used in other Guides). As a corollary of this, considerable effort was required in determining the “granularity”, or level of detail to which the content should be tagged. Creating a customised language for the encoding, storage and display of data created new demands, as it was necessary to keep track of the variants in language and how they work. For these languages to work, those involved in the production/delivery of educational content had to implement the languages describing this content according to a common standard (which was not entirely bad, in that it compelled all of those involved to work towards adopting

common standards for the storage, presentation and dissemination of information, as I discuss in further detail below.)

XML publishing places necessary constraints on the authoring process to ensure that documents conform to a set structure. On the one hand, these restrictions can reduce the number of choices that needed to be considered in the construction of a document, so that authors could (in theory) concentrate on the content and structure rather than presentation (which is left to a style sheet). On the other hand, there was a valid concern that adopting an overly structured approach may constrain the scope within which authors could write.

Another problem encountered at the delivery stage was that XML was not recognised by some Web browsers. At the time, browsers such as Internet Explorer 6 displayed the Guide; however, older browsers such as Netscape 4 were unable to effectively process the XML, displaying the data as seemingly garbled strings of text. XML handling techniques were not user friendly (Kasdorf 2003, p. 91). Furthermore, as I mentioned in Part 2 earlier, a Web browser will reject an XML file if it is not well formed. For example, failure to put closing tags around a single piece of information resulted in the Guide appearing incorrectly in the end user's browser.

Accessibility was a critical issue. OCC users accessed the Web from a range of sites (home, classroom, staff room) via a range of browsers and modem connection speeds. For this reason, the content of the Guide needed to be fast; however, there were problems when Web users requested content from the Guide in a PDF format to print from work or home. FOP tended to slow the process of delivery. The process had to be refined for usability to increase the speed of PDF files produced "on the fly" (i.e. the PDF file was automatically spawned from the server) or retrieved from a cache.

Further technical challenges emerged when seeking to integrate XML with the broader IBO network. At the time of implementation (2002-2003), there was a lack of support on server, browser and development tool sides.

Benefits of XML

A key strength of using XML in a single source approach was to reduce the duplication of labour and data. Rather than modify the print and Web versions of the Guide separately, it was possible to edit and manipulate one source of content within that document. Specific tags and metadata were used to break the Guide up into modules or "learning objects" that were easier to update on a regular basis. Once the generic content was identified it was possible to automate its dissemination. For example, all Diploma Programme Guides featured the same standard introduction and copyright page. In future, these sections of the Guide would only need to be marked up once and automatically generated in all Guides for the both the Web and print as necessary. Using a single source in this way avoided unnecessary duplication of content. Because the basic content was the same for both versions, any amendments or future revisions only need to be made once. Storing all content in a single source means that there was no longer a need to store this content separately in different media. Different versions could be created as necessary through specific style sheets.

The systemic approach required for XML to work effectively compelled the creation of standards for the storage and presentation of content across the authoring, designing, formatting and deliver stages of content creation, which promoted a consistent approach.

Aside from longer term saving of human resources, the need to find more inexpensive and environmentally friendly ways of disseminating information was an important consideration and driver of Web based delivery. Given the large quantity of data on the OCC, which occupied space on the server and consumed energy, XML was an appealing system.

XML was particularly useful for divorcing the display from the content when delivering data via the Web. By looking at the relevant XML tags, a back end application could manipulate data in different ways (for example, see Figure 1 above) (Bradbury 2001). The print version was provided for teachers to reference as a linear outline of the syllabus, while the online version enabled teachers to access this content strategically using a variety of navigational paths, as well as access extra resources.

Another benefit of XML is “future proofing” the content of the Guide against technological and organisational change. Because XML is useful for storing information that will be presented in a variety of formats, it can be useful as an intermediate step while transforming from one format to another. During the last few years, XML has become widely adopted as data interchange format in areas such as digital publishing (Kasdorf 2003, p.86). XML can be converted to other formats relatively easily, so that even if XML becomes obsolete, the source file for the Guide could, in principle, be converted to the new standard for printing and/or the Web. Given that Web based technology changes so rapidly, this ability to future proof data storage and handling will undoubtedly be beneficial in the longer term.

Perhaps the greatest benefit arose from the conceptual and educational challenges arising from the need to rethink how the content of the Guide could be restructured for the Web. Devising the structure of the content for the OCC website compelled reflection on some of the basic principles of educational design for the Web. Educational needs drove technological change. The educational and design principles underlying this approach to the Guide will be familiar to practitioners of e-learning and Web design (Collis 1999; Krug 2000; Collis & Moonen 2001; Sorenson 2002). XML was implemented as a foundation to the development of online content as learning objects that can be reassembled and reused in a variety of ways. Implicit in this educational approach was a model of learning according to which the process of accessing and using teaching and learning resources on the Web was viewed as a key part of the process of knowledge formation itself. Essentially, the development and use of Web based resources was understood as a process in which data is converted to information, and information converted into knowledge. The presentation of the Guide was structured according to the principle that the *process* of interacting with the Subject Guide should help the teacher to obtain and make sense of the information she/he needs in both practical and conceptual ways.

For the Economics Guide, content producers were challenged to rethink the online structure and presentation of the Guide in terms of: (i) information hierarchy (i.e. how information is organised); (ii) navigation (i.e. how users work their way through content); and (iii) design and interface (i.e. how content looked online). Categorising and prioritising information within the Guide was a critical step to developing a “blueprint” for how it would be delivered online. This stage of development commenced with a logical breakdown of the Subject Guide into its constituent chapters, followed by an exploration of thematic links between and within those parts. Essentially, the Subject Guide was broken into learning objects that could be arranged and rearranged according to the needs of the user. For example, the area of “Development Economics” was a discrete section of printed version of the Economics Guide, but was mentioned throughout the Guide as a recurring theme as well. The intention was to enable the user to access all of the information about Development Economics directly from the single source in no more than three mouse clicks. But it was also necessary to limit the degree of ‘chunking’ of information as too many chunks became difficult to mark up and manage. These chunks were therefore prioritised according to: which ones should be given more prominence on screen; which information had to be seen by the user; and which could be optionally viewed. The structure for chunking information was identified to enable flexibility as to how these could be maintained by content providers, and accessed and displayed onscreen by users. Where certain categories were considered of equal importance, then the design of the user interface was critical to visually organising this information on screen efficiently and effectively without overloading the user.⁹

Consequently, the authors of the Economics Guide were invited to rethink how they could develop the structure of curriculum beyond a conventionally linear, static printed document; but also as a dynamic, organic body of learning objects that can be interconnected and presented in a variety of ways. Authors were invited to adopt a more spatial approach to curriculum development for the Web, which brought about fresh, new thinking and challenges. For example, in rethinking how the structure of Economics Subject Guide could be redeveloped for the Web, the introduction of XML compelled content authors to reflect on questions, such as: How can the authors of curriculum imagine content differently? What can be done online that is not done via other media, such as paper? How can the Web be used to extend/deepen understanding of the subject area? How are end users expected to use the content? Are they expected to upload their own resources, download resources (such as PDF versions of the Guide)? Engage problem solving activities or simply glean information?

⁹ The rule of navigation, for example, was that the content for the online version of the Guide had to be structured in such a way that the user could access the necessary information within three clicks (unless the content require some form of extended problem solving or interpretative exercise).

Standard printed content that was previously designed in a static and conventionally linear “narrative” was explored as a non-linear space, whose structure was visualised as a three dimensional mezzanine in which content is dynamic, interrelated and shaped by the collective efforts of content producers and users, as opposed to a typical printed guide, which, when imagined as a space, had a linear structure consisting of a corridor (the contents page of the Guide) leading to several separate and distinct rooms (chapters within the Guide).

Development of the online version of the Economics Guide also stimulated reflection upon how the Guide could be more interactive. The Web version of the Guide enabled a high degree of user control over the exploration of the subject area online. Rather than bombard teachers with information that may be irrelevant to their interests and needs, the online version of the Guide was flexibly structured to be customisable by end users, to empower them to seek and discover information of most interest and relevance. Throughout the process of the development of the Subject Guide, it was important that the authors and other producers of educational content remained focused on the core benefits of XML, to drive the printing and online delivery of content to add value to existing conventional practices (e.g. in print).

As a result, content became more “Web friendly” and more easily managed and updated. By maintaining only one source of content, the Subject Guide could be updated regularly and potentially be more responsive directly to users needs. In summary, the process of reflection stimulated by a single source approach using XML promoted a more organic approach to content development, shifting the use of the Web from a means of disseminating informational data, to thinking about online educational content as a virtual space of living, breathing knowledge.

Conclusion

The pilot of single source publishing approach using XML was considered a success, and the Subject Guide for Economics went live on the OCC on 28 February 2003, albeit in a modestly basic form. By April 2003, the basic framework was in place to expand development of the framework to publishing and repurposing other curriculum materials and resources (e.g. Maths Guides). The DTD written for the Economics Guide could be applied to a range of similar documents (i.e. other Subject Guides) featuring a similar structure.

Since then, the range of XML tools, vocabularies, parsers and published DTDs has grown. Consequently, the use of XML by the IBO has changed considerably. YAWC was abandoned due to instability and usability issues. New tools and applications, have since been trialled, such as *upCast* to streamline the use of MS Word to author XML documents and *oxygen* to edit XML. Refinement and improvements in the technology have impacted upon workflows.

XML was adopted because it is extensible, platform independent and supports internationalisation and localisation by enabling the same content to be manipulated according to the different needs of users. As XML uses text, any platform that can handle text should, in principle, be able to handle XML. The XML standard is free to access and enables the possibility for powerful data management, sophisticated use of metadata, reducing duplication of human resources, displaying content in different ways and enhanced online interactivity of content across different sites of delivery. There were also educational benefits consistent with the use of educational content in database driven approaches to e-learning.

The most visible benefits of adopting XML in the manner described in this paper were long term rather immediate. This approach would appear to work best on content that is intended for delivery on a large scale, and the process was not easy! Earlier I cited Kasdorf’s (2003, p. 87) assertion that XML could be used on projects as small as a single book or journal; however, I would suggest to educators that XML may be inappropriate for ‘one off’ or small projects. Furthermore, XML will not work if generic standards are not adopted for the mass dissemination of content. The true value of the Economics Guide pilot will be determined once the processes of data storage, handling and thinking about content as learning objects are applied across other subject areas within the IBO. The use of this kind of system requires a close collaboration between educational authors, designers and ICT technical support. In the case of the Economics Guide, this collaborative process of rethinking how content was stored and delivered was every bit as rewarding as the outcome or any other technical benefits XML had to offer.

References

- Apache Software Foundation (2004). FOP. *The Apache XML Site*. <http://xml.apache.org/fop/>
- Biz/ed (2004). Biz/ed Website. [viewed 20 July 2004] <http://www.bized.ac.uk/>
- Bradbury, D. (2001). How XML will change your life. *Create Magazine*, December.
- Collis, B. (1999). Designing for differences: Cultural issues in the design of WWW-based course-support sites. *British Journal of Educational Technology*, 30(3), 201-215.
- Collis, B. & Moonen, J. (2001). *Flexible Learning in a Digital World*. London: Kogan Page.
- Harold, E.R & Means, W.S. (2002). *XML in a Nutshell: A Desktop Quick Reference*. O'Reilly.
- IBO (International Baccalaureate Organisation) (2003a). *Economics Guide*. International Baccalaureate Organisation.
- IBO (International Baccalaureate Organisation) (2003b). *Economics Guide*. Online Curriculum Centre. International Baccalaureate Organisation [viewed 10 Apr 2003] <http://online.ibo.org/>
- IBO (International Baccalaureate Organisation) (2004). *International Baccalaureate Organisation*. [viewed 15 Ju 2004] <http://www.ibo.org/>
- Kasdorf, W. E. (2003). Markup: XML and related technologies. In W. E. Kasdorf (Ed), *The Columbia Guide to Digital Publishing*. Columbia University Press.
- Krug, S. (2000). *Don't Make Me Think: Common Sense Approach to Web Usability*. New Riders Publishing.
- Merceica, P. (2001). Digital publishing: The current state of play. In B. Cope and D. Mason (Eds), *Creator to Consumer in a Digital Age: Australian Book Production in Transition*. Australia: Common Ground Publishing.
- Sorenson, N. (2002). Six criteria for evaluating your web site. *The Editorial Eye*, January, p. 8.
- W3C (2003). *World Wide Web Consortium (W3C)*. [viewed 20 Jul 2004] <http://www.w3.org/XML/Schema>
- W3Schools (2004). *XML Tutorial*. [viewed 28 Sep 2004] <http://www.w3schools.com/default.asp>

Acknowledgement and disclaimer

I managed the Online Curriculum Centre based in the United Kingdom from 2001 to 2003, during which time I oversaw the Centre's redevelopment. I wish to gratefully acknowledge the support and input of staff of the International Baccalaureate Organisation in the development of many of the ideas which are reflected in this discussion. In particular, I wish to acknowledge the practical insight provided by Andrew Ashton and John Knowles, with whom I worked during the implementation of XML at the IBO. Please note, however, that the opinions in this paper are the author's alone and are not representative of the IBO.

Using extensible mark-up language for the single source delivery of teaching resources via print and the Web: A practical example

Dr Lucas Walsh, Post-doctoral Research Fellow of New Media, School of Communications and Multimedia, Edith Cowan University. Email: l.walsh@ecu.edu.au

Please cite as: Walsh, L. (2004). Using extensible mark up language for the single source delivery of teaching resources via print and the Web: A practical example. In R. Atkinson, C. McBeath, D. Jonas-Dwyer & R. Phillips (Eds), *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (pp. 913-923). Perth, 5-8 December. <http://www.ascilite.org.au/conferences/perth04/procs/walsh.html>

Copyright © 2004 Dr Lucas Walsh

The author assigns to ASCILITE and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive licence to ASCILITE to publish this document on the ASCILITE web site (including any mirror or archival sites that may be developed) and in printed form within the ASCILITE 2004 Conference Proceedings. Any other usage is prohibited without the express permission of the author.