

Title: **XML opens doors**

Date: **June 2007**

XML opens doors

Developing a sizeable on-line distance programme will involve an institution in the creation, and subsequent management, of a significant volume of learning materials. These materials may take many forms, but there will be a significant amount of textual content employed. How should this content be structured, held and managed?

There are many tools that might do the job. For example millions of documents every day are written using Microsoft Word. Millions of users can't be wrong, so we could choose Word to manage our content. But can it meet our programme needs? For an effective delivery that suits the needs of our students we will probably need to deliver to:

- Paper, so that there is always a working base that every student can access and, anyway, students like books;
- **PDF**, so that students can have all the content on their laptops;
- **HTML**, so that we can populate our virtual learning environments (**VLEs**).

Word can do all of that as it has a "Save as .." feature that offers all of these file formats.

Alternatively, we could also author our content directly in HTML, as this would let us develop directly in the format used by our VLE - though we may need to keep a separate Word copy of everything as well as this HTML version. That's OK, provided that we keep everything in step.

Or we could be smart and keep a single source master of everything that:

- let's us produce to all the formats we need;
- stops duplication of effort; and
- is independent of tools and VLEs ... forever.

There is really only one choice if this route is taken, namely the wholesale adoption of XML.

What is XML?

XML stands for **eXtensible Mark-up Language**.

A mark-up language allows structure in a document to be tagged with some meaning, in a manner similar to structured use of common word processors. Like Microsoft Word, XML can be used to describe the *structure* of a document, but it is much more flexible and extensible and, indeed, not limited to text (for example vector images described in the **SVG** standard make use of the expressiveness of XML).

XML is an international standard, and a subset of **SGML** (ISO 8879), a mark-up language that has been used since 1986 to define the structure, and hold the content, of many types of documents. Though originally developed to be simpler than SGML, and targeted for web-based documents, XML can be used to describe a vast array of document structures (read learning materials), including those for print.

HTML – the language of the web page – is an *application* of XML, but a fairly simple one. XML can actually be used to create an infinite number of applications, of varying complexity, but it is generally the case that applications are designed for specific needs – from technical manuals for aircraft, to the specification of historical or news events, through to computer messaging. The popularity of HTML was based partly on its relative ease of use, but mainly because of its ability to display a wide range of documents – via a web browser – which could be linked to other documents.

However this popularity has led to demands for it to offer more, and this has exposed the limitations of HTML. It is a fixed application for the World Wide Web, not extensible, quite restrictive, and weak in its ability to describe the actual content of a document.

XML overcomes the inherent limitations of HTML, while providing many enhanced capabilities. Applications of XML make use of a ‘tag set’, just like HTML, but it is a *meta-language* – a language used to define other languages – rather than a single instance. So, while XML is actually used to define presentational applications such as HTML, there are many other instances that are more expressive. **DocBook** and **DITA** are popular instances for structured documents, and **IMS QTI** is a *de facto* standard for test and question mark-up for on-line assessments.

In an XML document, the tag names are designed to convey the meaning of the data they contain (cases, questions, answers). This is in contrast to HTML tags which are fixed (paragraph, heading, link) and therefore say little about the meaning of their content.

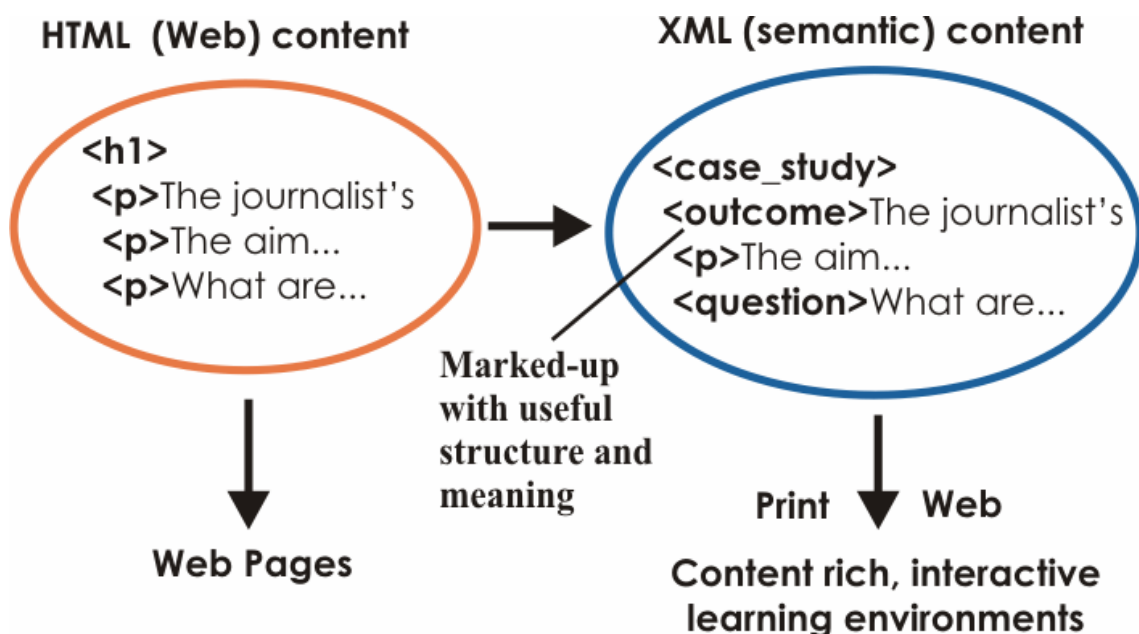


Illustration of the use of meaningful XML tag names which enable more interactive use in virtual learning environments. Case studies and questions can be made interactive, giving students a better on-line learning experience

So XML can be used to *semantically* tag documents in a way that adds meaning to their structure. This helps to improve the use and management of information, the publishing of it, and its exchange.

“We do XML”

Having extolled the virtues of XML, do be warned! XML is dumb – its usefulness depends on how it is applied and used. Any claim of “we do XML” should always be tempered by *how* it is being used.

CAPDM uses XML (most commonly with DocBook as the defining document type definition – **DTD**) as a standard for the capture, management and publishing of all client content. Used within a managed repository, and with a policy of single-source mastering, this offers protection for the life of the content, and a flexibility with respect to the range of final learning environments and media formats that can be generated and used.

It is important to recognise that content longevity and reusability comes from *structure and semantics*, and **not** from the *file format* that the content might be saved in. Using the **Save as... XML** feature in Microsoft Word really just changes the file format and nothing else. This may allow you to claim that “we do XML” but it is not the same as using XML in a meaningful and principled manner.

Using XML also frees an organisation from the use of specific - tools though it is important, and useful, to be using the ‘right’ tools for each stage of content development. For example, word processors *are* undoubtedly good for creative authoring and maybe visual editing, but they are not good for content mastering or for generating professional quality published output.

If you are serious about your content, why not use a serious toolset? You wouldn't keep your financial database in a set of flat text files; so why keep your course materials in a set of word processor files?

In XML creation and editing, a visual editor like Word will only ever be able to take you part of the way toward structure and semantics, because Word will neither constrain the edits you make, nor allow you to create and manipulate the extra metadata which makes your information more useful.

Conclusion

The printed technical manuals for a Boeing 747 are larger and heavier than the aircraft itself and represent a substantial body of knowledge that must be always kept up to date. The content domain for the UK's largest distance learning MBA programme at [Edinburgh Business School](#) numbers 20+ million words in 4 different languages, and 290+ separate course components, released in print and on-line formats.

If you think of all the processes you will want to apply to the content for your programmes over their lifetime – creation, editing, updating, high-quality printing, high-quality web output, re-structuring, archiving – then there is really only one *lingua franca* for all of those processes, and that is structured, semantic XML.

Glossary

This paper has introduced a number of acronyms and technical terms which deserve some explanation. *Wikipedia* (www.wikipedia.org) provides a useful and fuller explanation of each.

- **SGML**: a meta-language in which one can define mark-up languages for documents, originally designed to enable the sharing of machine-readable documents in large projects. It has also been used extensively in the printing and publishing industries, but its complexity has prevented its widespread application for small-scale general-purpose use.
- **DocBook**: a mark-up language for technical documentation, currently maintained by the DocBook Technical Committee at OASIS and available in both SGML and XML forms, as a DTD (Document Type Definition).
- **DITA**: the *Darwin Information Typing Architecture* is an XML-based, extensible architecture for authoring, producing, and delivering technical information, divided into small, self-contained content topics that can be reused in different deliverables.
- **IMS QTI**: the *IMS Global Learning Consortium* is a **non-profit** standards organization concerned with establishing interoperability for learning systems and **learning content** and the enterprise integration of these capabilities. Their mission is to "support the adoption and use of learning technology worldwide". Their main activity is to develop specifications, like QTI, which might eventually be adopted as standards.
- **QTI**: an IMS specification – the Question and Test Interoperability Specification – which defines an XML language for interchanging questions and assessments between different systems.

"CAPDM provides a range of professional services that help learning providers to develop successful businesses in education. Visit us on-line at www.capdm.com for more information."