Glossary of E-Book and XML Terms

The full version of this glossary is available on the AAUP Wiki,, http://aaupwiki.princeton.edu, under Electronic Publishing → Glossary of Terms. Please feel free to make corrections and suggestions.

ancestor: An *element* that encloses the *parent* of the current *element*.

app: Short for "application" or program. Formerly used by techies to refer to any program, today app more often refers to programs developed for mobile *device OS*es. A book app or e-book app is an e-book packaged as an application in order to provide capabilities (animation, interactivity) not available in current e-book readers. In some cases an e-book app is produced simply to provide another sales channel, and the app version provides no more content or capability than the e-book.

application file: The source file in which a book is typeset. This has typically been a proprietary format, such as Quark or InDesign.

attribute: Variable assigned within an *element* to alter how it is processed. Zero or more attributes may be placed within an opening *tag* as name="value" pairs, e.g., `<div type="acknowledgments">I wish to thank everyone.</div>`

catalog: A file that maps generic addresses to local directories.

character entity: A code placed in *XML* text to refer to a special character, such as a list bullet, check mark, or diacritic that is not easily represented in the *character set* of the document. See also: *parameter entity*.

character set: An *XML* document must use only one character set, and that set should be declared at the beginning of the document. In fact, to step outside the context of XML for a moment, a single text file of any type may only use one character set. The most basic set is ASCII, which includes only 95 printable characters. Another common set is called Latin-1, which includes 191 printable characters and can fully represent 26 languages. *Unicode* covers most languages. UTF-8 is a flavor of Unicode.

child: An *element* enclosed by the current element.

chunk: To split a large document into smaller parts such as chapters, appendices, articles, etc.

CSS: Cascading Style Sheets. CSS gives a web browser instructions on how to format and position *HTML element*s. It is not required, but is a very efficient way to apply consistent styles, such as font and font size, across several pages or documents.

customization layer: A stylesheet to be processed with existing standard stylesheets in order to supercede certain formatting instructions with custom instructions.

DAM: Digital Asset Management. Refers to the process of storing files in a repository optimized for archiving, retrieval, and search.

declaration: May refer to a *doctype* declaration, which is placed at the beginning of an xml document, or an *entity* declaration, which resides in a specfication. One can also declare a *namespace*.

descendants: Elements enclosed by the child *element*s of the current element.

device: Computer, cell phone, tablet, netbook, notebook, or dedicated ebook reader.

DocBook: A DTD maintained by OASIS (Organization for the Advancement of Structured Information Standards). A glossary specific to DocBook can be found here: http://www.sagehill.net/docbookxsl/glossary.html.

DOCTYPE: A *declaration* at the beginning of an *XML* document that tells you the *DTD* to be used and the *element* that will serve as root.

DOI: Digital Object Identifier. A unique id registered with and maintained by a *linking service* such as CrossRef (http://www.crossref.org). A DOI may point to any part of a digital object, but for e-books it typically references the text at the chapter or paragraph level.

DRM: Digital Rights Management. Software that prescribes and enforces the viewing time, prevents copying of source files, or otherwise limits what the user or purchaser can do with a digital object.

DTBook: DAISY or Digital Talking Book. A DTD maintained by the Daisy Consortium (http://www.daisy.org). DTBoook was originally developed for visually impaired or otherwise handicapped readers. It is now a part of the *ePub* standard (content of an ePub file may be *XHTML* or DTBook.)

DTD: Document Type Definition. A detailed specification for encoding and producing documents with *XML*. The DTD specifies which *element*s are valid, how they can be used, and to some extent, their customary meaning within the document.

ecosystem: Refers to a combination of software, hardware, and purchasing options that play well together. Commonly seen are the "Nook ecosystem" or the "Google ecosystem." For example, the "Kindle ecosystem" refers to Amazon's website, the Kindle reading device, and Kindle software that allows you to view Kindle books on other devices. The word seems to connote symbiosis and interaction, as well as the essentially closed nature of these systems.

element: Part of an *XML* document enclosed by an opening *tag* `<tei>` and a closing tag `</tei>`. Elements can be nested within each other. The root *element* is the first element of the document, and whose tags enclose the entire document, excluding declarations at the very beginning..

entity: An organization, company, or individual associated with the document, generally referred to in the *DOCTYPE* declaration (DocBook version 5) or *DTD*. See also: *character entity*.

ePub: A standard e-book format, supported by the International Digital Publishing Forum (IDPF http://www.idpf.org). From the IDPF site: "'.epub' allows publishers to produce and send a single digital publication file through distribution and offers consumers interoperability between software/hardware for unencrypted reflowable digital books and other publications."

HTML: HyperText Markup Language. Web pages are formatted with *HTML*. See also: *XHTML*.

linking service: A company that indexes and helps clients maintain *DOI*s.

metadata: Data that describes other data. For example, metadata can be placed in the header of an *HTML* or *XML* file, describing the file's contents and structure. Metadata is often formatted as *XML*, and there are metadata standards, such as *Dublin Core*.

Mobipocket: A suite of tools for packaging and reading e-books. The file format generated by Mobipocket Creator is .prc. These files can also be read natively on the Kindle. However, Mobipocket Reader can display many common file formats.

namespace: A standard specification that lists all of the rules governing *tag*s, attributes, etc. The *attribute* xmlns is used to declare a namespace. It points to the *URI* of a file (usually on the web) that provides the specification. Assigning a namespace to an *element* (generally the root *element*, and therefore the whole document) allows a *validation* tool to accurately determine what syntax is allowed in the document markup.

ONIX: An *XML* standard for the electronic exchange of book product *metadata*, maintained by EDItEUR (http://www.editeur.org).

open access: A publishing model that provides full online access, free of charge, to the publication. Revenue to cover costs is provided by institutional support, grants, author fees, and other means. A list of open access journals is maintained here: http://www.doaj.org.

open source: A software development model in which the source code for the software is posted online for anyone to analyze, test, and improve. Given a sufficiently active community of professional programmers, bugs and security problems are discovered and fixed more quickly than in proprietary software. Linux is an example of open source software.

OS: Operating System. Examples are Windows XP, Mac OS X, Android, and Symbian OS. All *device*s have an OS, but some are less well known, or are unique to a particular device. Among mobile devices, Google's Android OS runs the Barnes and Noble Nook, the Motorola Droid, and many others. The OS on the iPhone is referred to as the "iPhone OS." The Kindle OS is actually a modified version of Linux.

parent: An *element* that encloses the current *element*.

PDF: Portable Document Format. Developed by Adobe, and formerly proprietary, PDF is now an open standard.

platform: A software environment, usually proprietary, for providing a service or product such as e-books. Not to be confused with *OS* or *ecosystem*. For example, one can speak of the "Kindle platform" and thereby refer to all the software that displays kindle books, whether it is on the Kindle Reader, and iPod, or a PC. By contrast, the Kindle "ecosystem" would also include your personal account on Amazon.com and the arrangements Amazon has made with publishers.

processor: An *XML* processor is a program that transforms XML into *HTML*, XML, or formatted output, such as *PDF* or postscript. Some common processors are Saxon, xsltproc, and Xalan. However, a program like <oXygen/> (http://www.oxygenxml.com) will let you choose which processor to use. Also, recent web browsers can perform some of the basic functions of a processor to convert XML to *HTML*, as long as they are given a stylesheet. InDesign also functions to some extent as an XML processor.

schema: The successor to *DTD*s. Like DTDs, schemas describe the valid *elements* and their legal usage in a document. Schemas are written in *XML*, support *namespace*s, are more extensible and easier to use and transform than DTDs. For the purpose of discussing a

workflow, schema and DTD are sometimes used interchangeably.

stylesheet: A file that contains formatting instructions to be carried out a file that contains marked-up text. *XML* stylesheets may contain expressions in *XSLT, XPath,* and/or *XSL-FO.*

tag: Tags are the basis of *XML* markup. An opening *tag* consists of a left angle bracket followed by a word with no spaces, followed by a right angle bracket: `<tag>`. A closing tag is exactly the same, except that the word inside the brackets is preceded by a right slash:`</tag>`. Tags surround content `<tag>`content`</tag>` to create an *element*. Tags are case-sensitive.

TEI: A set of guidelines maintained by the Text Encoding Initiative (http://www.tei-c.org) for the encoding of machine readable texts. An overview of TEI can be found here: http://tei.oucs.ox.ac.uk/GettingStarted/html/in.html.

template: There are several types of templates in *XML* processing, both in the generic sense of an incomplete framework or starting point, and in very specific senses. For example, stylesheets may contain template directives, which are snippets of *XSL* that explain how to transform XML when a particular pattern is encountered.

validation: All programs that use *XML* perform some type of validation. A common web browser can tell you if the XML is well-formed, that is, if opening and closing *tag*s match and there are no illegal characters. Real validation checks that the XML in the document conforms to the *DOCTYPE* declared at the start of the document and that no undeclared *namespace*s are used.

Unicode: Intended to be the holy grail of *character set*s, covering thousands of characters and most language scripts. UTF-8 is a Unicode character set that is expected in XML unless otherwise specified. Often it is specified anyway for clarity.

URI: A file location, expressed as a network path. The path is usually to a file on the Internet, but it may be to a shared local resource, such as an intranet fileserver, or even to a file on your hard drive. Stands for Uniform Resource Identifier.

XHTML: Extensible *HTML*. Essentially the same as HTML, but it conforms to *XML* syntax. XHTML and HTML documents look very similar at first glance, but an XHTML document can be parsed by an XML processor, not just a web browser.

XML: Very simply put, *XML* is a way of marking up all the text in a document so that it can be styled and transformed reliably by a computer program. It stands for eXtensible Markup Language. Just as with a spoken language, we use the term to refer to the whole language itself, as well as samples. So we say something is "in XML" or "that's XML" interchangeably. XML itself is not proprietary, is maintained by a standards organization, and is used in wide variety of data interchange.

XPath: In stylesheets, XPath is the language used when specifying where a transformation should be applied. It describes a hierarchy of *tag*s or *element*s.

XQuery: A language for querying XML data.

XSL: eXtensible Stylesheet Language. Actually a family of languages: XSLT, XSL-FO, XPath. A language for manipulating *XML.*

XSLT: eXstensible Stylesheet Language Transformations. XSLT specifies how to style *XML* text.