

Introduction to DocBook

Norman Walsh

Sun Microsystems, Inc.

About DocBook

- **DocBook is a schema maintained by the DocBook Technical Committee of OASIS**
- **Available as an SGML or XML DTD, RELAX NG Grammar, or W3C XML Schema**
- **Particularly well suited to books and papers about computer hardware and software (though by no means limited to these applications)**
- **About 10 years old (it will be 13 on 10 November 2005)**

About Norman Walsh

- **Member of the Java Web Technologies and Standards group at Sun Microsystems, Inc.**
- **Chair of the DocBook TC**
- **Active participant in web standards at W3C (XML Core, XSLT, TAG) and OASIS (DocBook, Entity Resolution, RELAX NG)**
- **Specification lead for JSR 206: *Java API for XML Processing***
- **Long-time markup geek**

A Little Background

Structured Documentation

Benefits

Challenges

Storing Structured Documentation

XML or SGML or ...

OASIS

Evolution

Structured Documentation

- **Semantic rather than presentational**
- **Components have *identifiable* structure**
- **ASCII and Word (without templates) are not structured**
- **HTML and Word are somewhat structured**
- **DocBook is strictly structured**

Benefits

- **Multiple presentations from the same source (print, online, help, etc.)**
- **Documentation reuse**
- **Authors no longer have to worry about presentation**
- **Opportunities for improved authoring interfaces**

Challenges

Technical

- **Relatively sophisticated processing required for presentation**
- **Document reuse requires careful management**
- **Users benefit from special authoring tools**

Non-Technical

- **Writing reusable documentation is different**
- **Authoring with structure is different**

Storing Structured Documentation

- **XML is the natural system for storing structured documentation**
- **XML can be used to develop different vocabularies**
- **DocBook is an XML vocabulary designed for computer documentation**

XML or SGML or ...

- **DocBook has historically been SGML**
- **DocBook 4 is supported in XML and SGML DTDs**
- **DocBook 5 will be XML primarily with hooks to allow enabling SGML-only features**
- **There are XML Schema, RELAX, and TREX Schemas for DocBook, but none are official at this time**

OASIS

- **OASIS: The Organization for the Advancement of Structured Information Standards**
A non-profit, international consortium that creates interoperable industry specifications based on public standards such as XML and SGML. OASIS members include organizations and individuals who provide, use and specialize in implementing the technologies that make these standards work in practice.

OASIS (Continued)

- **DocBook is the work product of an OASIS Technical Committee**

Evolution

- **DocBook is *stable***
- **Backwards incompatible changes can only occur at full version revisions (5.0, 6.0, etc.)**
- **Backwards incompatible changes have to be announced a full version before they are implemented**
- **Minor revisions (3.1, 4.1, 4.1.2) are always backwards compatible**
- **Starting with DocBook V5.0, the normative schema will be expressed in RELAX NG**

DocBook Markup

Element Classes

"Information Pool" Elements

Inlines

Inline Examples

Inline Categories

Linking

Paragraphs

Examples, Figures, Tables, ...

...

Element Classes

- **There are two main classes of elements in DocBook**
- **“Hierarchy” elements provide gross structure**
- **“Information Pool” elements provide prose markup**
- **The information pool could be reused in a new hierarchy**
- **Conversely, the hierarchy could be preserved with a new technical vocabulary**

"Information Pool" Elements

- **Inlines (publishing, linking, markup, user interfaces, programming, operating systems, ...)**
- **Examples, figures, tables, and equations**
- **Graphics (media objects)**
- **“Verbatim” (program listings, screens, ...)**
- **Admonitions (caution, warning, note, ...)**
- **Lists (ordered, itemized, simple, ...)**

Inlines

- There are roughly *100* inline *elements*:
- they identify commands (`ls`),
- code fragments (`x := 4`),
- dates (`2005-10-06`), etc.
- The `phrase` element is a general purpose wrapper.

Inline Examples

```
<para>There are roughly <emphasis>100  
</emphasis> inline  
<glossterm baseform="element">ele-  
ments</glossterm>: they identify  
commands (<command>ls</command>),  
code fragments (<code>x := 4</code>),  
dates (<date>2005-10-06</date>), etc.  
The <tag>phrase</tag> element is a  
general purpose  
<phrase>wrapper</phrase>.</para>
```

Inline Categories

- **Technical** (`package`, `termdef`, ...)
- **Error related** (`errorcode`, `errorname`, ...)
- **Programming** (`function`, `varname`, ...)
- **Products** (`productname`, `trademark`, ...)
- **Operating system** (`envar`, `filename`, ...)
- **Markup related** (`tag`, `token`, `literal`, ...)
- **Bibliographic** (`citation`, `author`, ...)
- **Publishing related** (`acronym`, `footnote`, ...)
- **Graphic** (`inlinemediainobject`)

Inline Categories (Continued)

- **Keyboard related** (`keycap`, `shortcut`, ...)
- **Indexing** (`indexterm`)
- **GUI related** (`guiicon`, `guibutton`, ...)
- **Links** (`link`, `xref`, `olink`, `anchor`)

Linking

- **DocBook uses ID/IDREF linking**

- > `<link linkend="someid">hot text</link>`

- > `<xref linkend="someid" />`

- **DocBook V5.0 adds XLink**

- > `<link xlink:href="someURI">hot
text</link>`

- > `<command xlink:href="#someid">ls</com-
mand>`

- > **Experimental support for link bases**

Paragraphs

- The DocBook paragraph element is `para`.
- For paragraphs with titles, there's `formalpara`.
- In DocBook, `para` can contain “block” elements (tables, figures, procedures, etc.). The `simplpara` element can only contain in-lines.

Examples, Figures, Tables, ...

- **DocBook has `example`, `figure`, `table`, and `equation`. These elements are “formal” and are expected to have a title.**
- **If you don't want a title, use `informalexample`, `informalfigure`, `informaltable`, and `informalequation`.**
- **Tables come in two flavors:**
 - > **CALS tables and**
 - > **HTML tables**

Graphics

- **Media objects (mediaobject):**
 - > **Images (imageobject),**
 - > **Video (videoobject),**
 - > **Audio (audioobject), and**
 - > **Text (textobject)**

```
<mediaobject><imageobject>  
<imagedata fileref="graphics/db2html.png" />  
</imageobject><textobject>  
<phrase>Converting DocBook with XSLT</phrase>  
</textobject></mediaobject>
```

Case Study: MediaObjects

```
<mediaobject>
<imageobject>
  <imagedata fileref="emc2.svg" />
</imageobject>
<imageobject>
  <imagedata fileref="emc2.eps" format="EPS" />
</imageobject>
<textobject>
  <para>Energy is equal to mass times the speed
of light squared.</para>
</textobject>
<textobject>
  <phrase>E=mc2</phrase>
</textobject>
</mediaobject>
```

Verbatim environments

- **Program listings:** `programlisting`
- **Screen shots:** `screen` (for command-line interfaces) and `screenshot` (for graphical UIs)
- **Literal layouts:** `literallayout`
- **Addresses:** `address`.

The `programlisting` and `screen` elements are generally monospaced; `literallayout` and `address` are usually in the same font as the body text.

Admonitions

- **note, tip, important, caution, and warning**

Note

This is a note.

```
<note>  
<para>This is a note.</para>  
</note>
```

Lists

- `itemizedlist` and `orderedlist`,
- `variablelist`, and
- `simplelist`

```
<itemizedlist>  
<listitem><para><tag>itemizedlist</tag> and  
<tag>orderedlist</tag>,  
</para></listitem>  
<listitem><para><tag>variablelist</tag>, and  
</para></listitem>  
<listitem><para><tag>simplelist</tag>  
</para></listitem>  
</itemizedlist>
```

Definition or “Variable” Lists

varlistentry Wraps each term (or terms)
and the definition.

term Wraps each term, there
may be more than one.

listitem Wraps the definition.

```
<variablelist>
```

```
<varlistentry>
```

```
<term><tag>varlistentry</tag></term>
```

```
<listitem><para>Wraps each term (or terms) and...
```

```
</para></listitem>
```

```
</varlistentry>
```

```
...
```

Special Purpose Markup

- **Function and command synopses**
- **Object-oriented programming classes, interfaces, methods, etc.**
- **Sets of messages**
- **EBNF diagrams**
- **MathML and SVG**

Case Study: Function Synopsis

```
<funcsynopsis>
<funcsynopsisinfo>
#include <pwd.h>
</funcsynopsisinfo>
<funcprototype>
  <funcdef>struct passwd *<function>getpwnam</function>
  <paramdef>const char * <parameter>name</parameter>
</funcprototype>
<funcprototype>
  <funcdef>struct passwd *<function>getpwuid</function>
  <paramdef>uid_t <parameter>uid</parameter></parameter>
</funcprototype>
</funcsynopsis>
```

"Hierarchy" Elements

- **Set and Book**
- **Part and Reference**
- **Preface, Chapter, Appendix, Bibliography, Glossary, Index**
- **Article**
- **Section, Sect1...Sect5, SimpleSect**
- **RefEntry**
- **RefSect1...RefSect3**

Case Study: A Book

This is the DocBook XML source for a book.

```
<book>
<bookinfo>
  <title>An Example Book</title>
  <author>
    <firstname>Norman</firstname>
    <surname>Walsh</surname>
  </author>
  <copyright>
    <year>2004</year>
    <holder>Sun Microsystems, Inc.</holder>
  </copyright>
  <contractnum>1234</contractnum>
```

Case Study: A Book (Continued)

```
<contractsponsor>Our Favorite Sponsor
</contractsponsor>
</bookinfo>
<preface><title>Introduction</title>
<para>...</para>
</preface>
<chapter><title>The First Chapter</title>
<para>...</para>
</chapter>
<!-- ... -->
<appendix><title>An Appendix</title>
<para>...</para>
</appendix>
</book>
```

Case Study: An Article

This is the DocBook XML source for an article.

```
<article>
<articleinfo>
  <title>An Example Article</title>
  <author>
    <firstname>Norman</firstname>
    <surname>Walsh</surname>
  </author>
  <copyright>
    <year>2004</year>
    <holder>Sun Microsystems, Inc.</holder>
  </copyright>
```

Case Study: An Article (Continued)

```
</articleinfo>  
<section><title>A Section</title>  
<para>...</para>  
</section>  
<appendix><title>An Appendix</title>  
<para>...</para>  
</appendix>  
</article>
```

Case Study: Reference Pages

This is the DocBook XML source for a reference page.

```
<refentry>
<refmeta>
<refentrytitle>getpwnam</refentrytitle>
<manvolnum>3</manvolnum>
</refmeta>

<refnamediv>
<refname>getpwnam</refname>
<refname>getpwuid</refname>
<refpurpose>get password file entry</refpurpose>
</refnamediv>
```

C.S.: Reference Pages (Continued)

```
<refsynopsisdiv><title>Synopsis</title>
```

```
<synopsis>
```

```
#include <pwd.h>
```

```
#include <sys/types.h>
```

```
struct passwd *getpwnam(const char * name);
```

```
struct passwd *getpwuid(uid_t uid);
```

```
</synopsis>
```

```
</refsynopsisdiv>
```

```
<refsect1><title>Description</title>
```

```
<para>The <function>getpwnam</function> function  
returns a pointer to a structure containing the
```

C.S.: Reference Pages (Continued)

broken out fields of a line from
<filename>/etc/passwd</filename> for
the entry that matches the user name
<parameter>name</parameter>.

</para>

<!--...-->

</refsect1>

<!--...-->

</refentry>

Tools

Applications

XSLT

Jade

Applications

- **Arbortext Epic**
- **oXygen**
- **XML Mind XML Editor**
- **Emacs and nXML mode**
- **Among others...**

See also <http://wiki.docbook.org/topic/DocBookAuthoringTools> and <http://wiki.docbook.org/topic/DocBookPublishingTools>.

XSLT

- **XSL Transformations, part of the Extensible Style Language from the W3C**
- **Many processors available (XSLTC, Saxon, Xalan, xsltproc, ...)**
- **Uses XML syntax and XPath as an expression language.**
- **Produces HTML, Formatting Objects, XML**
- **Formatting Objects can produce PDF (via FOP, RenderX, AntennaHouse, etc.)**

Jade

- **Processor for DSSSL (ISO/IEC 10179:1996 Document Style Semantics and Specification Language (DSSSL))**
- **Understands both XML and SGML source documents**
- **DSSSL uses Scheme (Lisp) as an expression language.**
- **Produces HTML, RTF, PostScript/PDF (via JadeTeX)**

Published Documentation

DocBook to ...

Profiling

Chunking

I18N Support

DocBook to ...

- **PDF with XSLT/XSL Formatting Objects**
- **HTML (XHTML, etc)**
- **HTML Help**
- **Java Help**
- **Unix “man” pages**
- **WordML (experimental)**

Profiling

- Use *effectivity* attributes to identify classes of content: `userlevel`, `security`, `os`, `version`, `condition`, ...
- Select a combination of values for publishing: for example, “topsecret” and “online” or “novice” and “windows” and “version5”.
- Content is filtered according to the profile.
- Result is processed to produce the output format of your choice.

Chunking

- **For online presentation, you can produce an entire document in a single HTML file or**
- **Create individual files at various levels: for example, one chunk per chapter or one chunk per top-level section.**

I18N Support

The DocBook stylesheets support 59 languages out of the box: Afrikaans, Albanian, Amharic, Arabic, Azerbaijani, Bangla, Basque, Bosnian, Bulgarian, Catalan, Chinese (Simplified), Chinese (Traditional), Croatian, Czech, Danish, Dutch, English, Estonian, Farsi, Finnish, French, German, Greek, Gujarati, Hebrew, Hindi, Hungarian, Indonesian, Irish, Italian, Japanese, Kannada, Korean, Latin, Lithuanian, Mongolian, Norwegian, Nynorsk, Oriya, Polish, Portuguese

I18N Support (Continued)

(Brazil), Portuguese, Punjabi, Romanian, Russian, Serbian in Cyrillic script, Serbian in Latin script, Slovak, Slovenian, Spanish, Swedish, Tagalog, Tamil, Thai, Turkish, Ukrainian, Vietnamese, Welsh, and Xhosa

Customization

The Role Attribute

Subsets

Extensions

Restricting Role on Emphasis (DTD)

Restricting Role on Emphasis (RELAX NG)

The Role Attribute

- All elements have a `role` attribute
- Stylesheets can key off of role values:
 - > `<literal>` vs.
 - > `<literal role="widgetSpec">`
-
- DocBook *never* specifies role values

Subsets

- **Subsets constrain DocBook**
- **All documents that conform to the subset also conform to the full schema**
- **Enumeration of attribute values**
- **Removing elements**
- **Constraining content models**
- **Doesn't usually require stylesheet/tool customization**

Extensions

- **Extensions extend DocBook**
- **Documents that conform to the extension may not conform to DocBook**
- **Adding new attributes or elements**
- **Extending content models**
- **Extensions can also remove elements**
- **Almost always requires stylesheet/tool customization**

Restricting Role on Emphasis (DTD)

```
<!ENTITY % emphasis.role.attrib
          role      (normal|emphasis)      "normal"
>
```

```
<!ENTITY % docbook PUBLIC "-//OASIS//DTD DocBook XML
          "http://www.oasis-open.org/docbook/xml/4.1
>
```

```
%docbook;
```

Restricting Role on Emphasis (RELAX NG)

```
namespace db = "http://docbook.org/ns/docbook"
default namespace = "http://docbook.org/ns/docbook"

include "docbook.rnc" {
    db.emphasis.role.attribute =
        attribute role { "normal" | "emphasis" }
}
```

Derived Schemas

SolBook: The Sun Documentation DTD

Simplified DocBook

Websites

Slides

SolBook: The Sun Documentation DTD

- **The source for docs.sun.com**
- **Restrictions to aid authoring and enforce style**

Simplified DocBook

- **Only supports articles**
- **Far fewer block elements**
- **Far fewer inlines**
- **About 100 tags vs about 400**

Websites

- **Uses DocBook information pool**
- **Replaces most of the hierarchy**
- **A website is a tree of nested web pages**
- **Stylesheets support both flat and tabular, two-column navigation**
- **See nwalsh.com for an example.**

Slides

- **Based on simplified DocBook**
- **Replaces article with a set of slides**
- **Slides can be divided into sections**
- **Stylesheets support HTML and PDF**
- **This presentation is generated from Slides source**

In conclusion

Q&A

Resources

Resources

- **DocBook Home Page (<http://www.oasis-open.org/docbook/>)**
- ***DocBook: The Definitive Guide* (<http://docbook.org/tdg/>). Norman Walsh and Leonard Muellner. O'Reilly & Associates, Inc. 1st Edition October 1999. ISBN 1-56592-580-7**
- ***DocBook XSL: The Complete Guide*. Bob Stayton. Sagehill Enterprises. 3rd Edition February 2005.**

Resources (Continued)

- **This presentation is online at <http://nw-alsh.com/docs/presentations/doctrain2005/>**