



DocBook

DocBook OSDC December 2004

by
Scott Penrose



© 2004 Scott Penrose - <http://linux.dd.com.au/>





DocBook - This talk

- ⇒ What is DocBook
- ⇒ Why DocBook
- ⇒ How to convert
- ⇒ What other formats can be used
- ⇒ Makefile to make your life easier
- ⇒ Generally an introduction !





DocBook - What

⇒ What is DocBook

- DocBook is a markup (originally SGML now XML)
- DocBook is a markup language for writing:
 - Documents
 - Books
 - Articles
 - Manuals
 - (Man pages)
 - ...





DocBook - What

⇒ What is DocBook

- Not presentational

- Levels of presentation: From HTML -> WP -> Postscript
- FO – Format Objects





DocBook - What

⇒ What is DocBook

● Markup the:

- Top level: Abstract, Authoblurb, Synopsis ... (dozens)
- Content... (hundreds)
 - Caution
 - Example
 - Figure
 - Graphic
 - Index
 - Code
 - Command line – required and optional params
 - Screenshot
 - Warning
 - Function Synopsis
- Meta Information... (a few)
 - Anchor, Highlights, Glossary, Index, Table of Content, ...





DocBook - What

⇒ What is it good for

- **Informational only – NO PRESENTATION**

- Huge reduction in consumption of time of writing
- BIGGER reduction in changing presentation in the future

- **Search-able and more**

- Present a page with Synopsis
- Imagine how easy it would be for Nat Torkington to get Publisher, Topic and more

- **Unlimited ability to change output**

- No need for a style guide for a book !
- Can change company, presentation or any other information in one central place.
- Table of content, Index, Table of figures
- Download example code
- Check syntax of commands

© 2004 Scott Penrose - <http://linux.dd.com.au/>





DocBook - Why

- Open standard for documentation for more than a decade (12 years...)
- SEE: <http://www.oasis-open.org/>
- Purely information format and contains no presentation
- This means you can (and many people do) convert from DocBook to MAN pages.
- If you are looking for a word processing (presentation) format, then go to the Open Office Document Standard.
- which is supported by OpenOffice at <http://www.oasis-open.org/>



Docbook - Why

⇒ Why ?

- Output to any format
- Fully searchable documentation database
- Existing tools (XML render, database etc)
- Force style based design (ie: you can't do text formatting, just choose the style and allow others to choose the rest) – forcing total separation between presentation and content.





DocBook - Why

- Very simple format, easy to get your head around
- XML/Text - choose your editor, grep, and all standard tools
- XML/Text - choose your programming language (often missed as an advantage of XML)
- Fully informational and no formatting.





DocBook - Why

- Technical documentation format used by myinternet.
- Not concerned with presentation.
- Ability to Print or PDF
- Include logos and company standards
- Allow changes to standard formats in one location
- Other output formats
- MAN pages for packaging
- Web pages
- Built in full search in Web portal





DocBook - Why

- Combinations
 - Vary the way the documents are combined
 - Single installation / configuration document for a particular software product
 - Combined configuration document for all products
- Sources (not just DocBook)
 - POD (Perl Modules etc)
 - APT (Almost Plain Text)
 - XXX (Other formats)
 - DocBook (sort of obvious one)





DocBook - Why

- **Converting**
 - Presentation markup formats such as TeX and HTML can also be converted to DocBook - giving you a very good common format for combining even some very esoteric formats.
 - Converting from presentation to informational is quite lossy but works usually enough to do something like a fraction of a chapter.
 - Other than headings and paragraphs, formatting is lost.
 - MS Word, HTML, Text, POD, APT, TeX and more...





DocBook - Why

- **Output Independence**
 - Apart from combinations, outputting the document is completely isolated. You can output from the same source document a highly structured PDF with letter head, HTML, Plain Text and even MAN Pages.
 - **Outputs include:**
 - HTML, XHTML, WML (single, chunked, any layout)
 - FO, PS, PDF
 - Text, TeX, JavaHelp
 - Man Pages, Info Pages
 - Presentations (multiple output types)





DocBook - Where is it Used

- Internally
 - Many organisations use DocBook internally
- Books
 - O'Reilly has a number of books in DocBook
- Computer Documentaiton (and Open-Source)
 - Xfree86
 - GNOME
 - KDE
 - FreeBSD
 - Linux
 - PHP
 - Debian
 - Blender
 - VIM





DocBook - Where is it Used

- Automatic production
 - From code (eg: GNOME and Linux Kernel)
- Embedded
 - Eg: in other code like XSLT
- Commercial Used
 - SUN, RedHat, Compaq, Rational, Apple...
- (list could go on for a while)





DocBook - History

- In the Beginning =
 - 1991
- The HaL and O'Reilly Era
 - 1991 to 1994
- The Davenport Era
 - 1994 to 1998
- The Oasis Era
 - 1998 to now





DocBook - Source Format

- **SGML DocBook**
 - SGML DocBook is often used (see Debian Documentation Project).
 - SGML is harder to maintain, parse etc. The same reasons XML was invented are why we don't use the SGML versions.
 - Fairly easy to convert from SGML to XML (with-in reason)
 - SEE: Makefile dbk->xdbk





DocBook - Source Format

- XML DocBook
 - SEE: Makefile xdbk -> *
 - Use anything that can edit or display XML
 - Text editors
 - Dedicated XML editors
 - WYSIWYG XML editors
 - Can use CSS or XSLT to convert or display in Mozilla

```
<?xml version="1.0"?>
<refentry id="Ml::config::database">
  <refmeta><refentrytitle>Ml::config::database</refentrytitle> <manvolnum>1mi</manvolnum></refmeta>
  <refnamediv><refname>/etc/mi/config/*</refname><refname>dbi_handle</refname>
    <refpurpose>Choose and configure the database used by MID</refpurpose>
  </refnamediv>
  <refsect1>
    <title>Description</title>
    <para>
```



DocBook - Source Format

- **APT = Almost Plain Text**
 - No effort to learn, looks exactly like plain text
 - Easy to email and share with others (no need to convert first)
 - Fantastic for CVS and Diffs
 - Downside = Not open source
 - SEE: <http://www.xmlmind.com/aptconvert.html>
 - SEE: Makefile apt->xdbk

This is the title of a section

This is a paragraph which contains several sentences. Sentence one. Sentence two. Sentence three. This paragraph is followed by a bulleted list.

* List item 1.

* List item 2. This *<word>* will be rendered in italic.
This other **<<word>>** will be rendered in bold.



DocBook - Source Format

- ⇒ POD = Plain Old Documentation
 - Almost as good as APT
 - But – open source
 - Embedded documentation in Perl





DocBook - Source Format

- Microsoft Word
- HTML
- POD - Perl Modules
- LaTeX
- RTF





DocBook - Internals

- DocBook Bits
 - DocBook (full set, very big)
 - Simplified (easier to support)
 - MathML
 - SVG





DocBook - File Type vs Extensions

- File identification
 - File Type, MIME Type, Magic Number
 - For convenience during this talk I am considering these the same thing.
 - MIME type are not necessarily connected to file type. For example a TSV file (Tab serperated data) as either text/plain or text/tab-separated-values - or even application/octet-stream.
 - MIME only has a very few types for XML (application/xml, text/xml) - this is actually completely inaccurate - more to come.





DocBook - File Type vs Extension

- File identification continued...
 - Extension
 - The obvious one - .doc, .txt, .xdbc
 - XML Type (schema, DOCTYPE, etc)
 - Need to effectively parse the XML doc to find the type
 - Very flexible, open and extensible
 - Hard to make work with traditional tools (Make, OO)





DocBook - File Type vs Extension

- Conclusion
 - Using '**xdbk**' to indicate 'XML DocBook'
 - '**dbk**' to indicate 'SGML DocBook'
 - '**apt**' to indicate APT (Almost Plain Text)
 - '**pm**' or '**pod**' to indicate POD
 - and the usual extensions





DocBook - Convert: * -> DocBook

- **Converting to DocBook**
 - The first part of the problem is to convert what you have to DocBook.
 - Once in DocBook you can combine documents from multiple source formats and do a single output.
- **APT**
 - 'aptconvert' allows you convert your APT file into DocBook

```
aptconvert output.xml input.apt  
mv output.xml output.xdbk
```





DocBook - Convert: * -> DocBook

- POD (Plain Old Documentation)
 - Pod2docbook
 - SEE: <http://search.cpan.org/~jaiv/Pod-DocBook-0.06/pod2docbook>
 - Pod2DocBook creates SGML

```
pod2docbook -infile=input.pod -outfile=output.dbk  
sgml2xml -xlower output.dbk > output.xdbk
```





DocBook - Convert: * -> DocBook

- MS Word
 - Ignoring all the file reading problems etc, this is obviously going to be very tricky, especially as it is mostly presentation information.
 - Word View - <http://wvware.sourceforge.net/>
 - Note that this takes three steps. Work on a direct convert template for wv is in progress.

```
wvHtml -targetdir=/tmp infile.doc outfile.html  
html2db outfile.html > outfile.dbk  
sgml2xml -xlower outfile.dbk > outfile.xdbk
```





DocBook - Output

- Why Output
 - Just kidding :-)
- DSSSL
 - SEE: <http://www.jclark.com/dsssl/> & <http://docbook.sourceforge.net/>
 - Been around for quite some time
 - Works with SGML
 - Written in LISP - Well... Scheme.
- XSLT
 - More active development now
 - More flexible
 - SEE: <http://docbook.sourceforge.net/>





DocBook - Output: XSLT

- XSLT = XML Style Language Translator
- SEE: <http://www.w3c.org/Style/XSL/>
- Engines
 - XSLT and how to make it happen is a whole other talk.
 - I use 'xsltproc'. Part of LibXSLT
 - SEE: <http://xmlsoft.org/XSLT/xsltproc.html>
- StyleSheets
 - I use stylesheets from <http://docbook.sourceforge.net/> with my own extensions.





DocBook - Output: XSLT

- Simple convert to HTML

```
xsltproc -o \  
output.html docbook-xsl/html/docbook.xsl \  
input.xdbk
```

- Chunked Output

```
xsltproc -o \  
outputdir/index.html \  
docbook-xsl/html/chunk.xsl \  
input.xdbk
```





DocBook - Output: XSLT

- Example of some options
 - TOC = Table of contents
 - TOC Section Depth = how many levels to include in this TOC output
 - Auto labeling. Every kind (sections, anything)
 - Generate INDEX
 - etc.





DocBook - Output: XDBK -> FO -> PDF

- FO = Format Object
 - To convert to PS / PDF I use FO (Format Object)
 - FO is the standard for presentation formatted objects (think PostScript)
 - 'output independent formatter'
- FOP = Format Object Processor from Apache
 - SEE: <http://xml.apache.org/fop/>

```
xsltproc -o output.fo docbook-xsl/fo/docbook.xsl input.xdbk  
fop.sh -fo output.fo -pdf output.pdf
```



DocBook

Output: XDBK -> FO -> PDF

- Example of some options
 - TOC etc from HTML
 - paper.type = A4
 - double.sided = 1
 - Generate index
 - Insert cross reference page numbers
 - Body font, footer font, font font font





DocBook - Makefile

- Use Makefile to handle converting
 - Automatically finds shortest path
 - Only builds what is needed
 - Fairly short
 - Remembers all those tricky commands
 - Does multiple steps (eg: sgml -> xml) that you often forget.
 - And all the other reasons to use a Makefile





DocBook - Makefile

⇒ A better make

- Since I wrote this make file I have learnt better techniques for rules
- These have not yet been applied
- Here are some examples
 - Instead of `.pod.dbk`: use `*.pod->*.dbk`





DocBook - Makefile

- Definitions
 - Put all your definitions at the top for easy maintenance.

```
# apt convert - Almost Plain Text - http://XXX  
APTCONVERT=aptconvert  
# XSL - XSLT for convert from DocBook to PS/HTML etc - http://docbook.sf.com/  
XSLT_PATH=/usr/local/docbook-xsl-1.50.0  
# FOP - Format Object P XXX - http://XXX  
FOP=/usr/local/fop-0.20.3/fop.sh  
# HTML2DB - Convert HTML to DocBook - http://XXX  
HTML2DB=/home/scottp/src/docbook/DocParse/html2db
```





DocBook - Makefile

- SUFFIX

- Each suffix must be defined for Make to automatically deal with them.
- I define everything I need in one place
- Think of Makefiles not as code but as rules

```
.SUFFIXES : .ps .pdf .xdbk .apt .html .lpr \  
.xv .pod .pm .dbk .xpdf .fo .acroread .tex \  
.doc .mozilla .svg .test .index .tar
```





DocBook - Makefile

- Your source convert (some examples)

```
.apt.xdbk:
```

```
  ${APTCONVERT} -toc -enc ASCII $(basename $@).temp.xml $<  
  mv $(basename $@).temp.xml $@
```

```
.pm.dbk:
```

```
  pod2docbook --infile=$< --outfile=$@
```

```
.doc.dbk:
```

```
  wvHtml --targetdir=/tmp $< tempfile.html  
  ${HTML2DB} /tmp/tempfile.html > $@
```



DocBook - Makefile

- DocBook output

```
.xdbk.html:
```

```
-mkdir $(basename $@)
```

```
xsltproc -o $(basename $@)/index.html ${XSLT_PATH}/mi/html.xsl $<
```

```
.xdbk.fo:
```

```
xsltproc -o $@ ${XSLT_PATH}/mi/fo.xsl $<
```




DocBook - Makefile

- Format Objects

```
.fo.ps:  
    ${FOP} -fo $< -ps $@  
  
.fo.pdf:  
    ${FOP} -fo $< -pdf $@  
  
.fo.svg:  
    ${FOP} -fo $< -svg $@
```



DocBook - Makefile

- Helpers

```
.dbk.xdbk:  
-sgml2xml -xlower $< > $@
```





DocBook - Makefile

- Viewers (shortcuts)

```
.pdf.xpdf:  
    xpdf $<
```

```
.pdf.acroread:  
    acroread $<
```

```
.pdf.mail:  
    mail -s
```

```
.ps.gv:  
    gv $<
```

```
.ps.lpr:  
    lpr $<
```

```
.ps.cups:  
    qtcups $<
```

```
.html.mozilla:  
    echo "file://`pwd`/$(dir $<)/index.html"  
    mozilla --remove 'openurl(file://`pwd`/$(basename $<)/index.html, new-window)
```



DocBook - Demonstrations

- APT = Demo_APT.apt
 - Make Demo_APT.xdbk
 - Make Demo_APT.html
 - Make Demo_APT.pdf
 - Make Demo_APT.acroread
 - Make Demo_APT.mozilla
- DOC = Demo_DOC.doc
 - Make Demo_DOC.mozilla
- myinternet potal





DocBook - Editors

- VIM :-)
- XML Editors
 - <http://www.xmlmind.com/xmleditor/>
 - DocBook and APT
 - Uses CSS so you can make it look like anything
 - WYSIWYG, including printing output
- Open Office :-)
 - Input/Output converter via XSLT
 - But... doesn't hide the features you don't have
 - Built in printing etc of course
 - Can generate invalid DocBook





DocBook - Other References

- docbook.org = Definitive guide h
<http://docbook.org/>
- DocBook Wiki -
<http://docbook.org/wiki/moin.cgi/>
- RefEntry = Man pages etc
<http://www.oreilly.com/catalog/docbook/chapter/book/refentry.html>
- Writing Documentation Using DocBook
<http://opensource.bureau-cornavin.com/crash-course/>
- A gentle guide to DocBook
<http://www-106.ibm.com/developerworks/library/l-docbk.html>
- And so many more – Try <http://www.google.com/search?q=docbook>

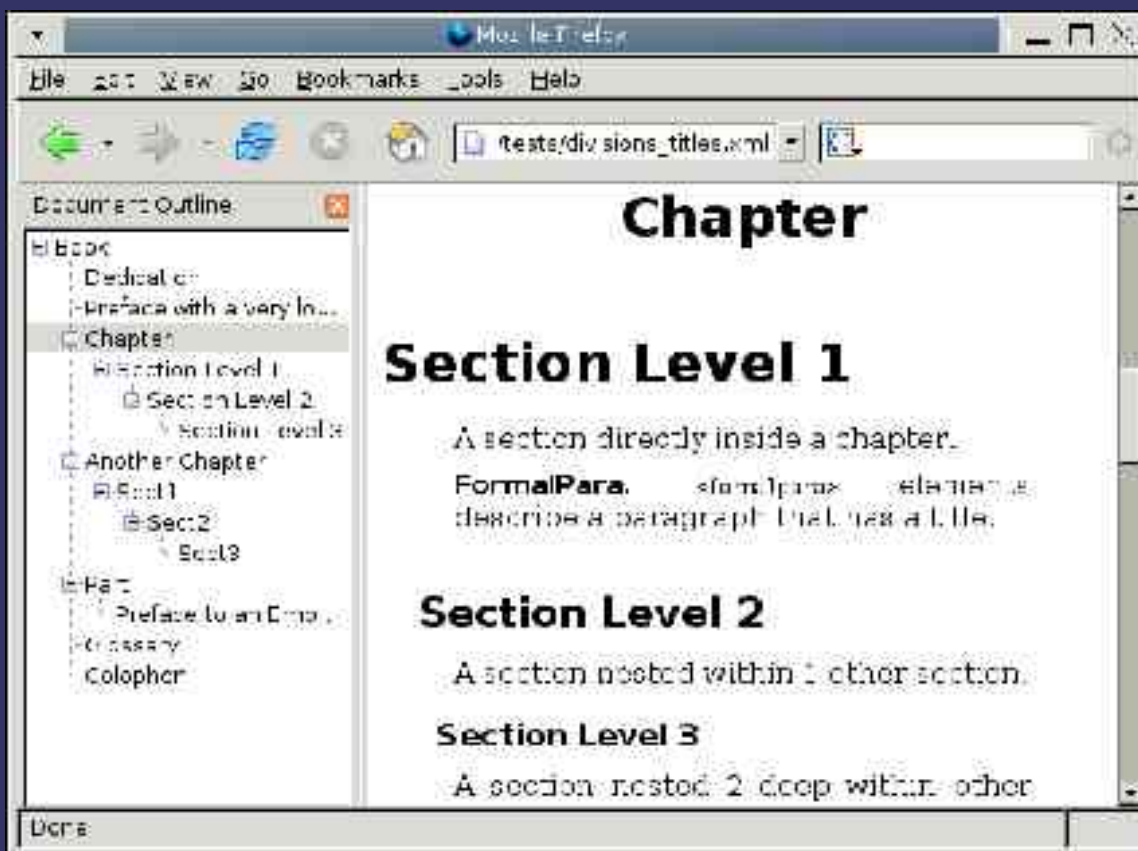




DocBook - Plugins

⇒ Mozilla (Outline Plugin)

- <http://www.badgers-in-foil.co.uk/projects/docbook-css/outliner/>





DocBook - Books

- ⇒ DocBook XSL: The Complete Guide by Bob Stayton and published by Sagehill
- ⇒ The definitive guide, written by Norman Walsh and Leonard Muellner and published by O'Reilly & Associates, Inc.





More users of DocBook

- ⇒ Debian Documentation Project
- ⇒ Linux Documentation Project
- ⇒ Blender 3D
- ⇒ The Vim Book
- ⇒ Bochs
- ⇒ OpenACS
- ⇒ And many many more
- ⇒ IBM and other commercial internal use not listed.





DocBook - The End

⇒ Questions ?

⇒ Content

- <http://linux.dd.com.au/narrative/osdc/docbook/>

