Migrating the Documentation Set to XML
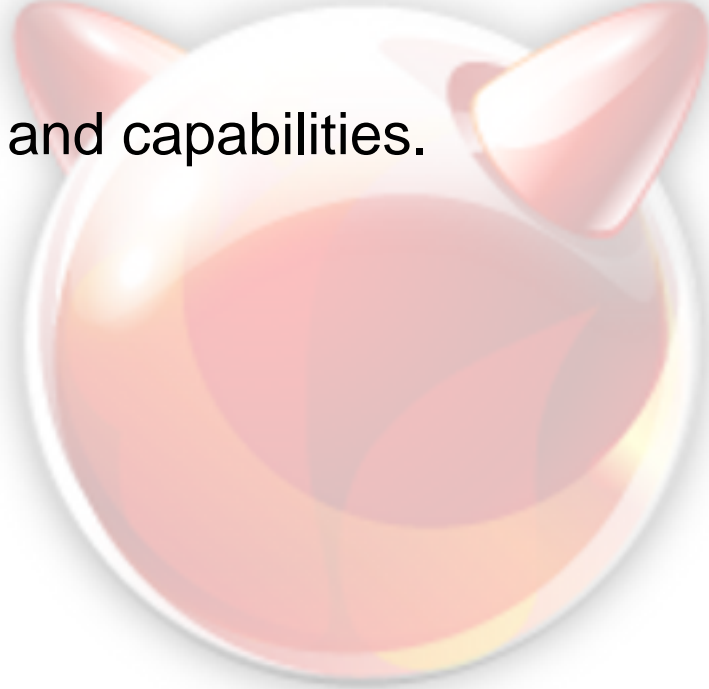
Gábor Kövesdán

`<gabor@kovesdan.org>`

20 Oct 2012

Motivation

- Better tooling, avoiding technology lock.
- Usability, rationalization of the infrastructure, easier learning curve.
- Modern outlook and capabilities.

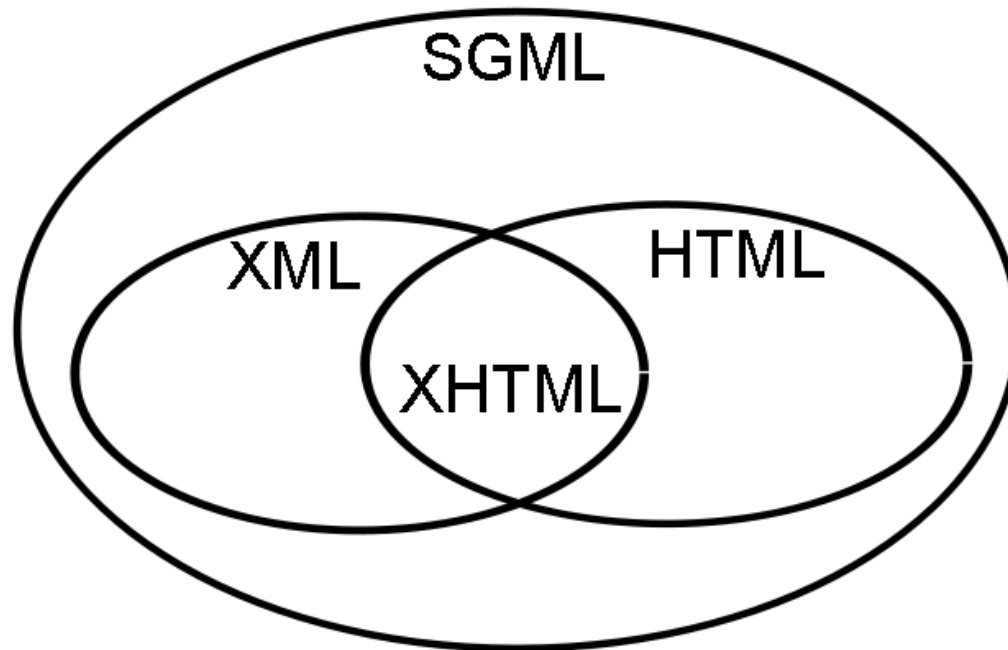# Migration plan

- Step 1: XML syntax and file names
- Step 2: Toolchain migration
- Step 3: DocBook 5.0

The SGML Family

XML versus SGML

- **K**eep **I**t **S**imple **S**tupid - Sometimes less is more...
- Simplicity lead to better tooling...
- ... and more and better complementary standards.
- SGML is practically dead.

Markup Changes

- No unclosed elements.

  E.g. `<br>` becomes `<br></br>` and with the shorthand nota-tion: `<br/>`

- No unquoted attributes.

- No attribute specification without attribute name.

  E.g. `<option selected>` becomes `<option selected="selected">`

- Names are case-sensitive.

- Dropped support for some shorthands.

  E.g. `<title>Introduction</title>` could be abbreviated as `<title>Introduction</>` or even `<title/Introduction/.`
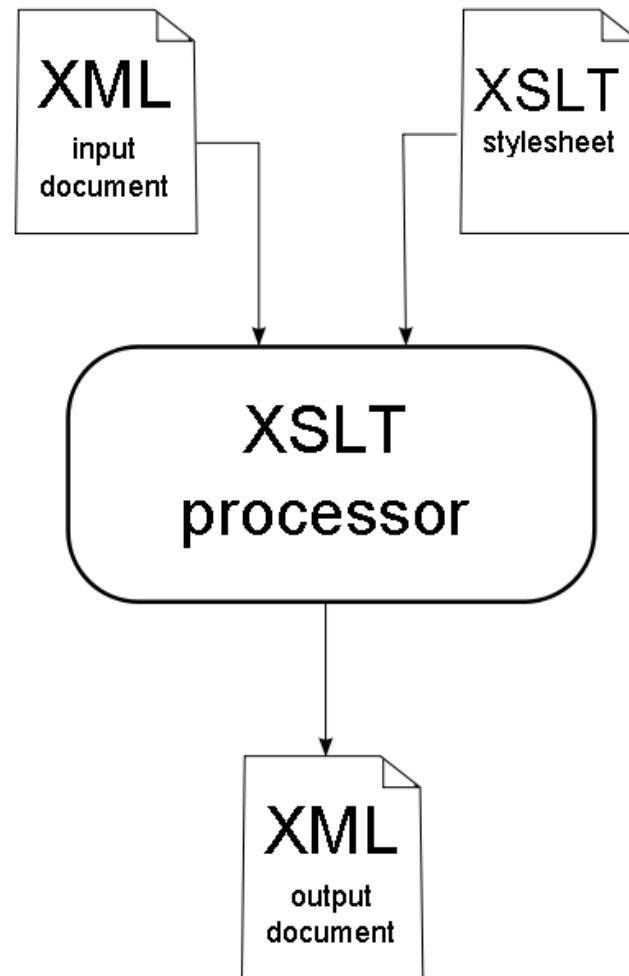
- Dropped support for some rarely used features, for example re-definition of the tag delimiters.

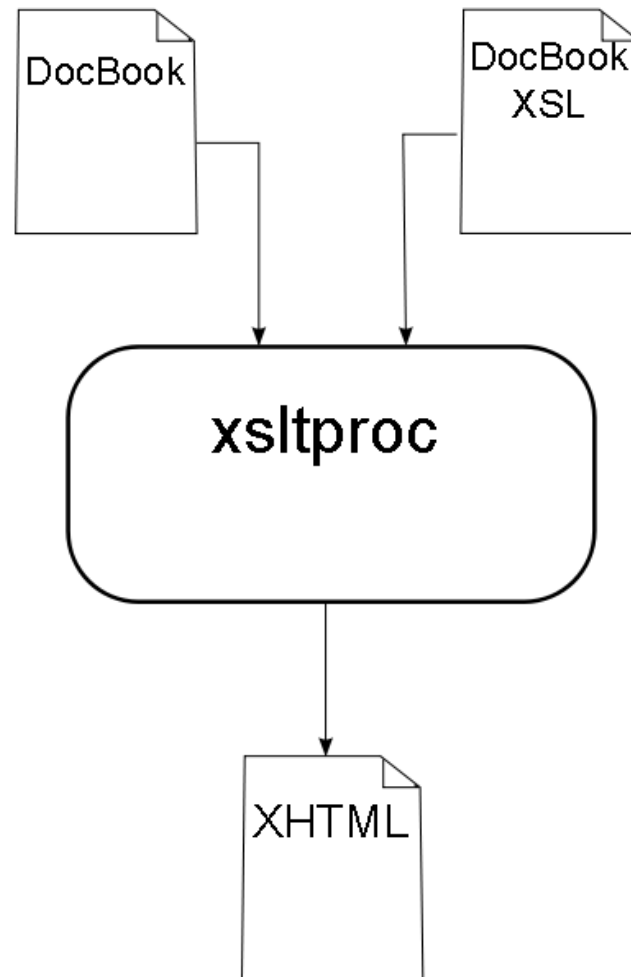# Not Strictly Related Improvements

- Dropped HTML tidy from webpages.
- Common XSLT fragments factored out.
- Sitemap, index sites and the multimedia section are now generated with XSLT.
- Merged entity sets.
- Untranslated entity sets pulled in from en_US.
- Avoid using CVS operations as part of the build.
- Removed emacs/PSGML comments.
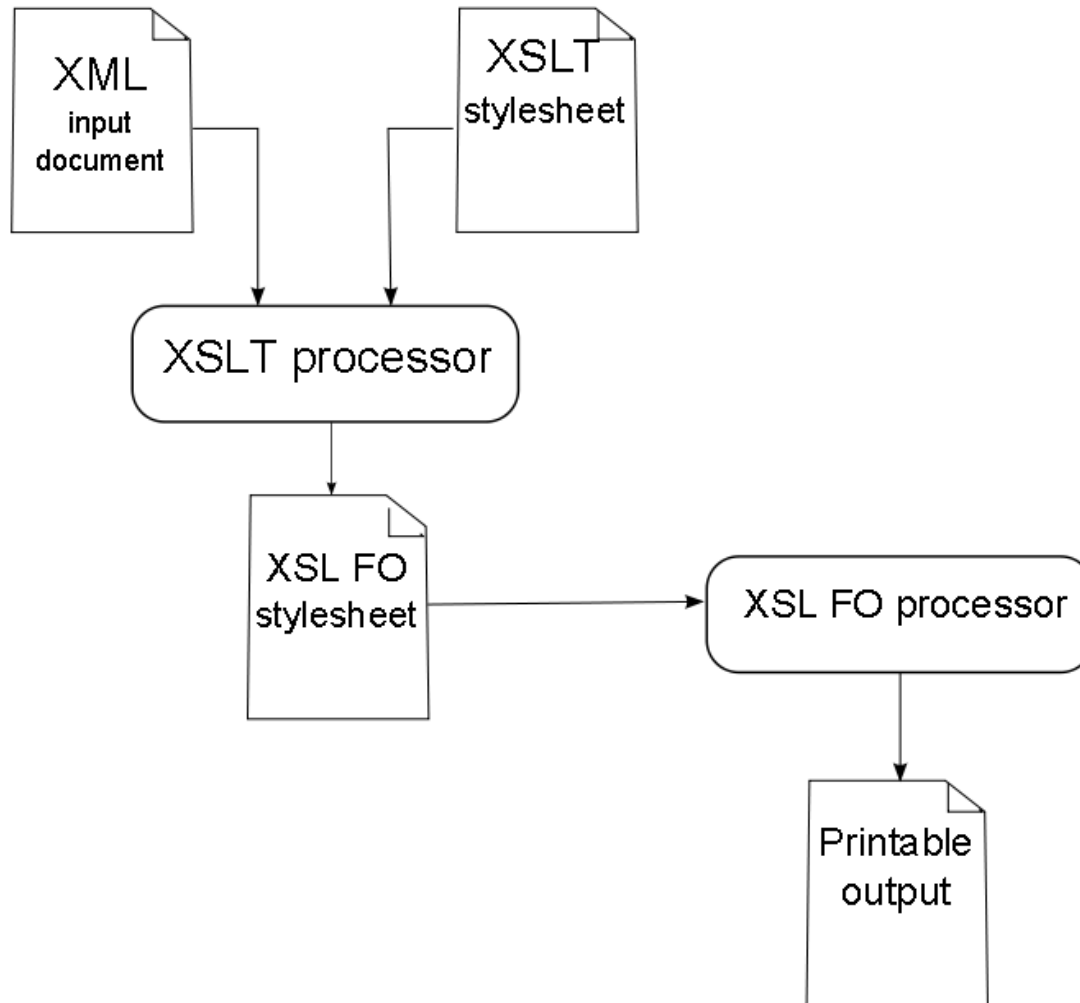- Removed web text size links.
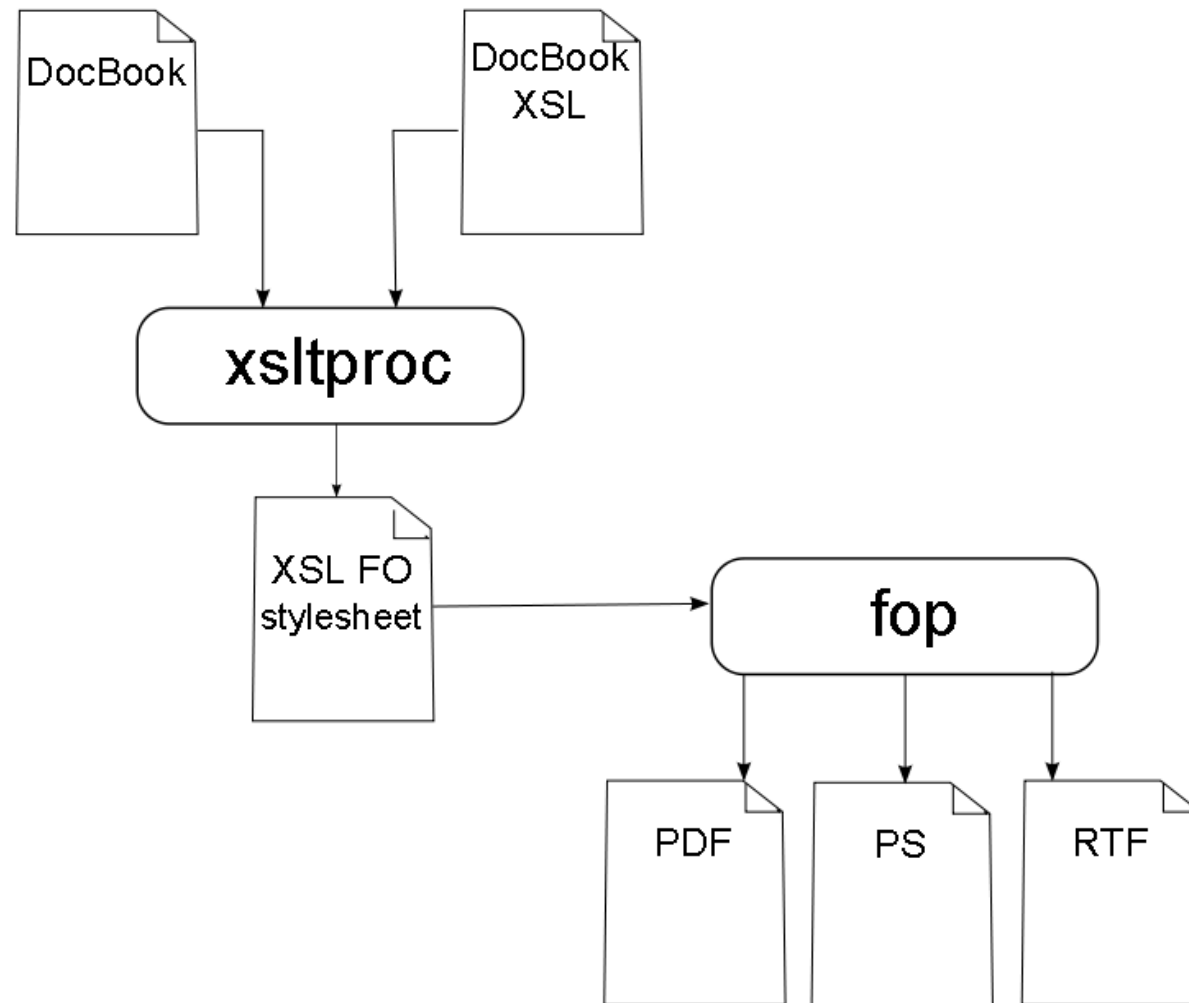- Dropped character entities.

Rendering Dataflow: XML output

Rendering Dataflow: XML output - XHTML

Rendering Dataflow: non-XML output

Rendering Dataflow: non-XML output

Demo

A currently generated PDF and a possible future one...

DocBook 5.0

- Simplified markup: info instead of articleinfo, bookinfo, sect1info, sect2info, ...

- Standard XLink linking and cross-references

- More advanced schema language: RELAX-NG

- Namespaces and easier extensibility.

Questions?

Thanks for listening!

Gábor Kövesdán

gabor@FreeBSD.org