

# XML and PDF

**Leonard Rosenthol**  
*Chief Technology Officer*  
*PDF Sages*



# XML

- HTML is a “markup language”
- XML is NOT a “markup language”
  - In spite of what the name says:  
“Extensible Markup Language”

# XML

- XML is a standardized way to create Markup Languages
  - many, many markup languages
- **Follow the XML rules – you get:**
  - Very well defined and documented rules
  - Common tools widely available
  - Common knowledge and understanding
  - Word search on any text portion
    - Selective by element context

# XML Markup Languages

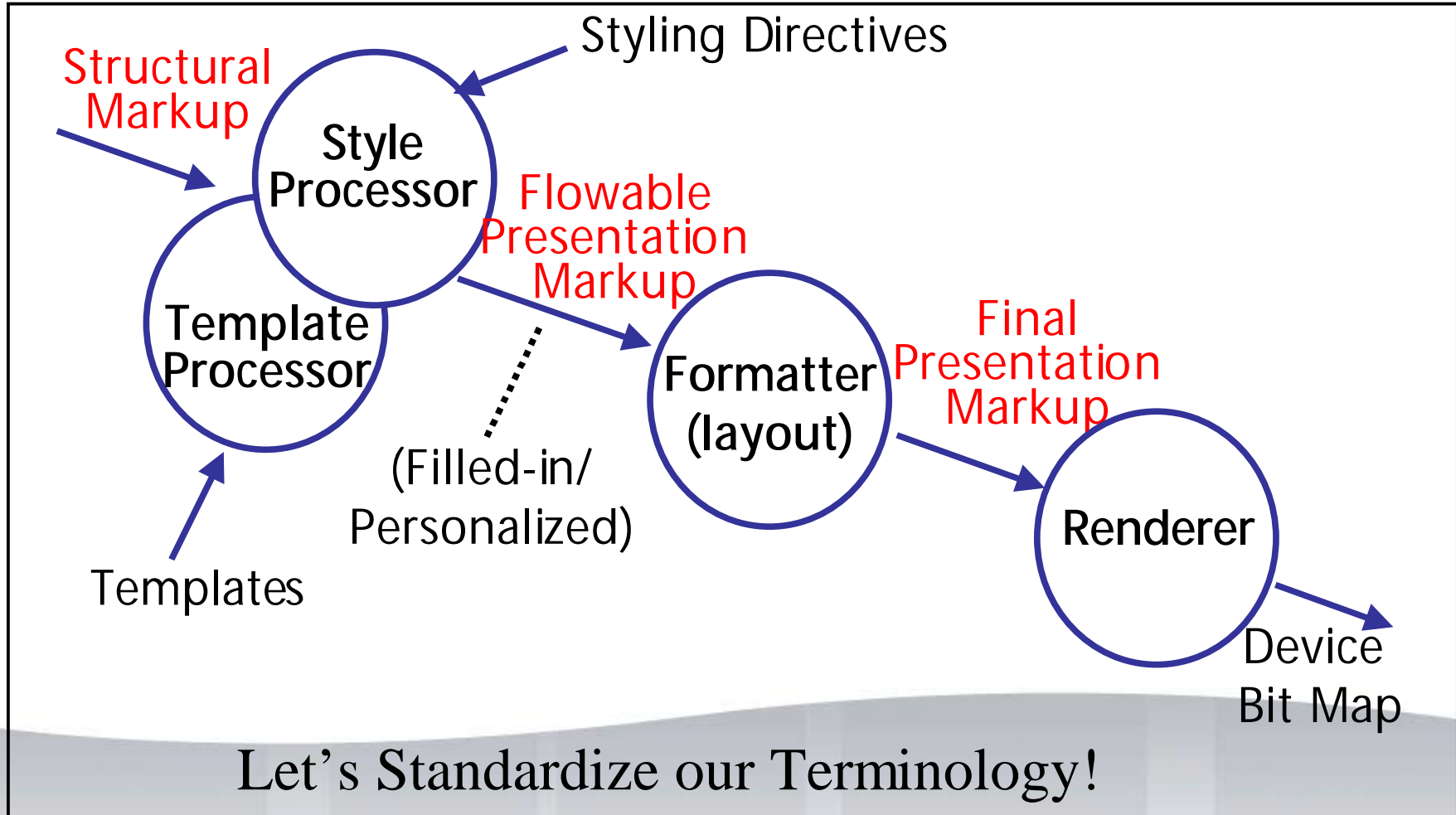
- Sample International Standards:
  - **SVG** – ***XML for Scalable Vector Graphics***
  - **OEB** – ***XML for Open Electronic Books***
  - **JDF** – ***XML for Job Definition Format (print)***
  - **XSL** – ***XML for Extensible Stylesheet Language***
  - **NewsML**– ***XML for News Stories***
- Or, check the “XML Catalog” at:  
<http://www.xml.org>

# XML for ...

- Must qualify what the XML is **for**
  - XML **for** *Business Cards*
  - XML **for** *HTML (XHTML)*
  - XML **for** *Print Job Definition Format (JDF)*
  - XML **for** *Scalable Vector Graphics (SVG)*
  - ...
- **OR**
  - XML **for** *Structure*
  - XML **for** *Flowable Presentation (XHTML)*
  - XML **for** *Final Presentation (SVG)*

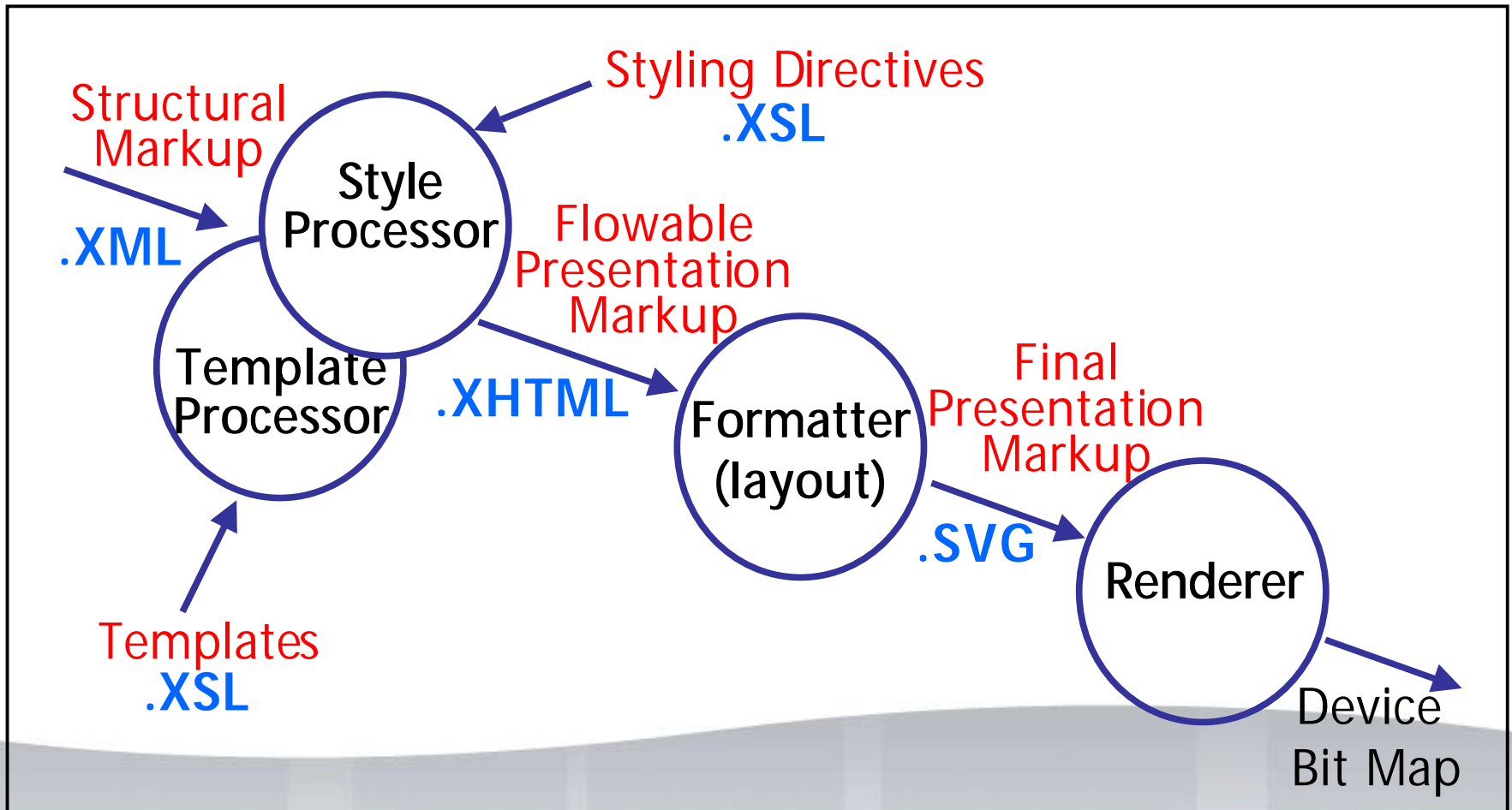
# Processing Markup (for Presentation)

- From Markup to Bit Maps

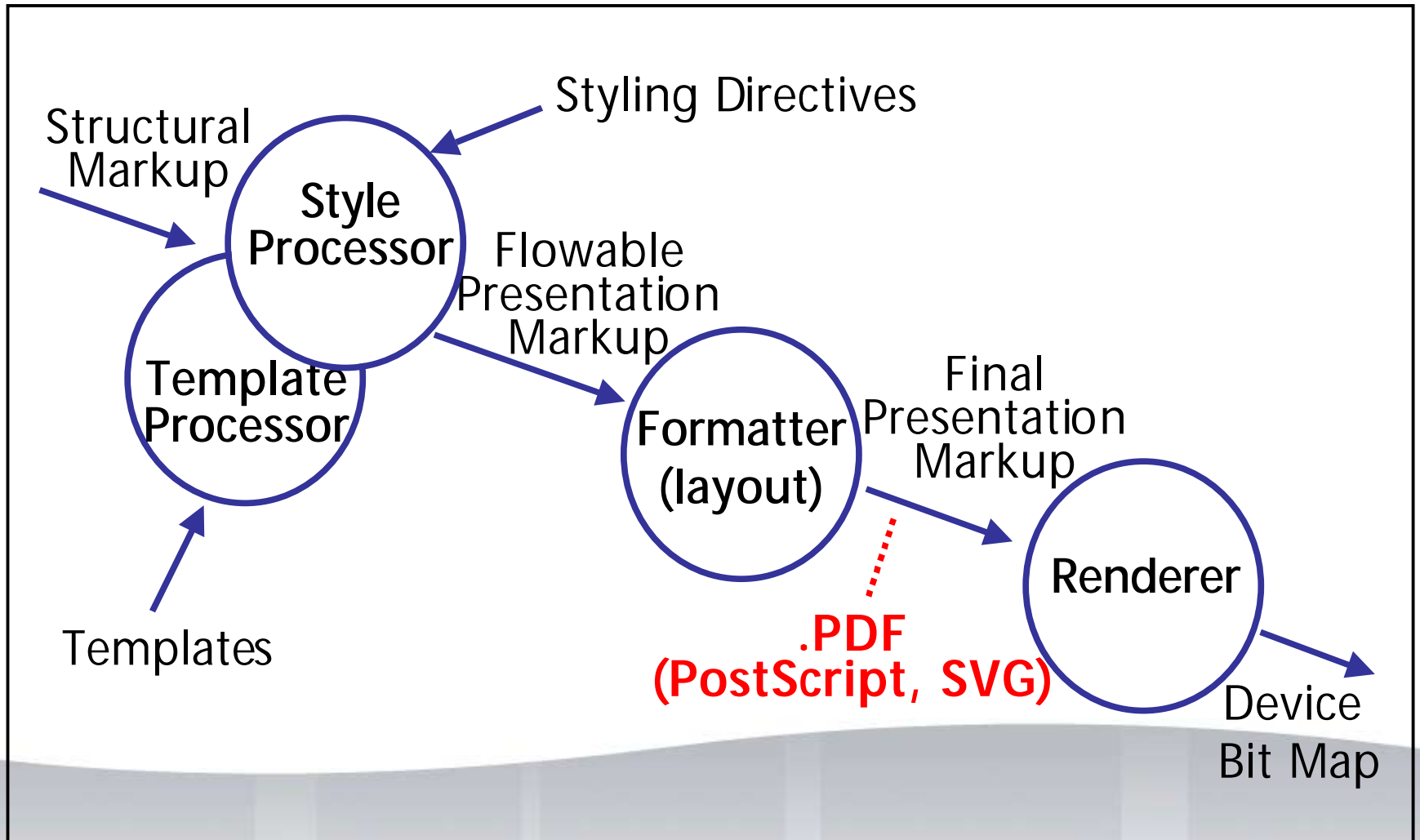


# XML Markup Languages

- XML markup languages exist for all **red** items



# PDF

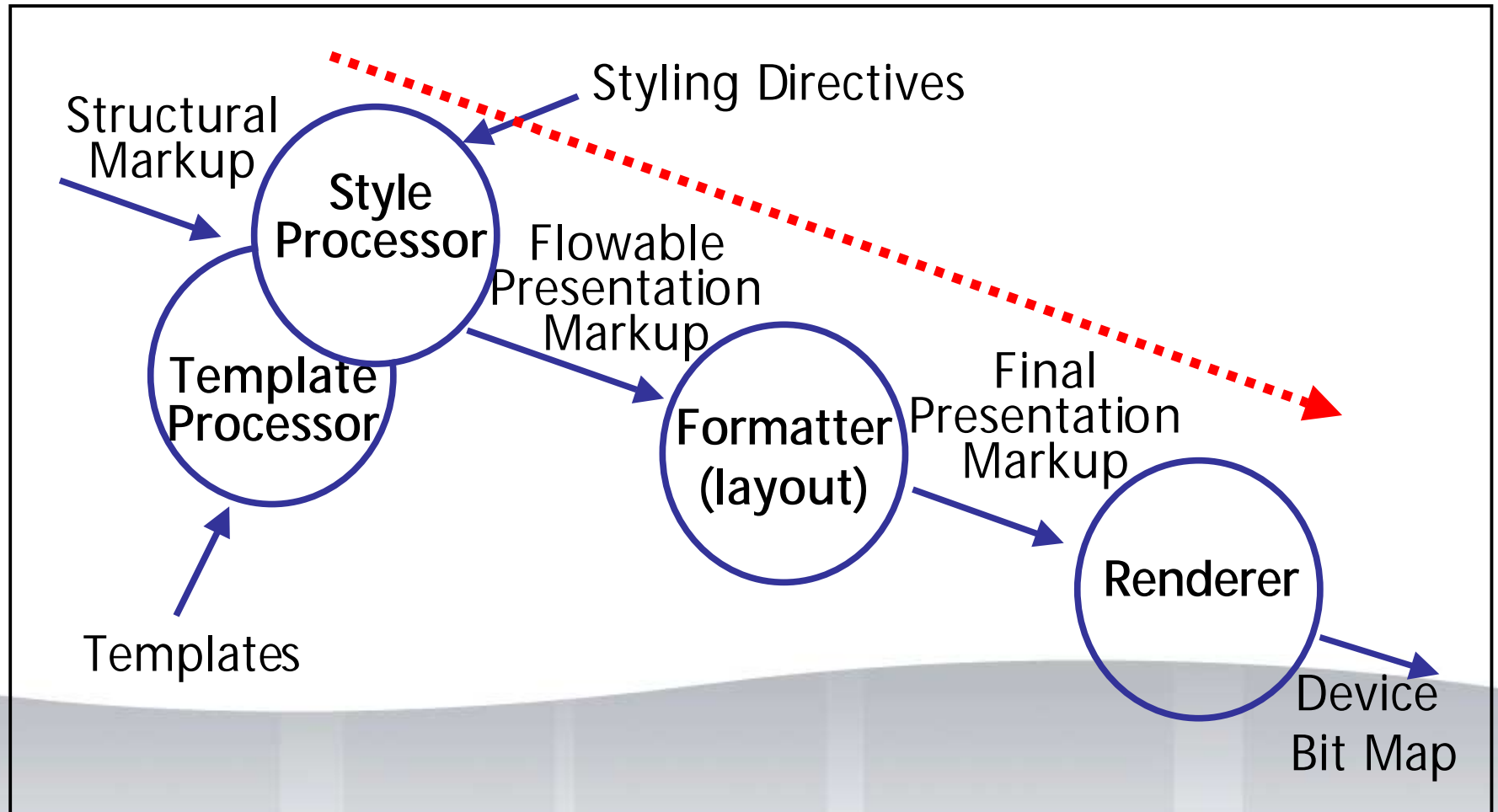


# But ...

- **I want one file to view**
  - on my laptop
  - on my PDA
  - on my E-book reader
  - on my cellphone
  
- **Must reflow text (at least)**

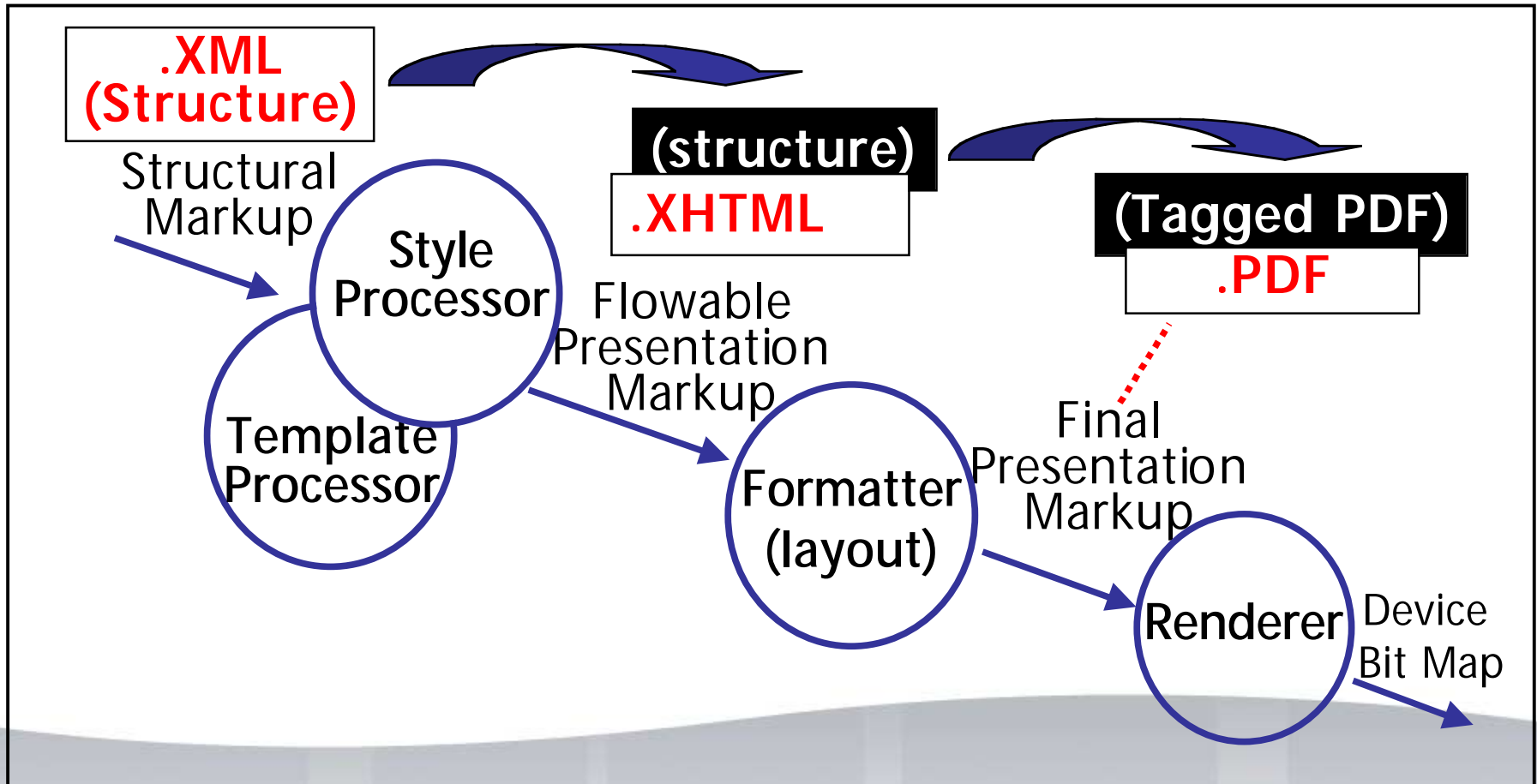
# With XML

- Redo processing (styling by rules)



# Reflow of Tagged PDF

- Structural information normally lost
- Add ability to save it



# Reflow of PDF

- Partially reprocess

