XML Inside PDF:
Digital Master for Document Structure

Chuck Myers
Technology Strategist

Tuesday, December 11, 2001

Agenda

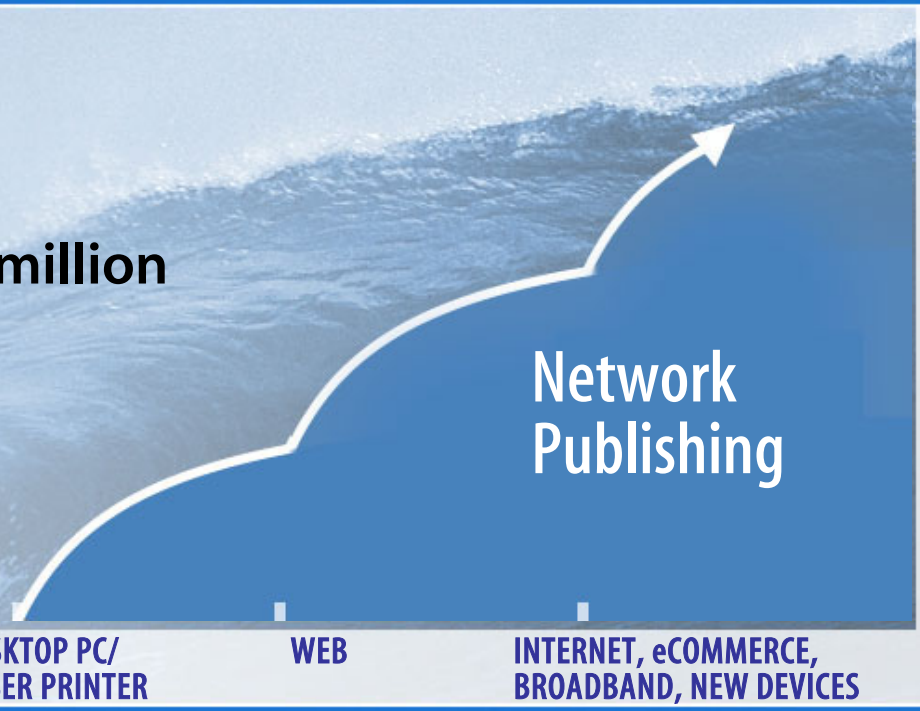
- XML/Publishing Overview
- XML/PDF Integration in Acrobat 5.0
- Resource Pointers

Network Publishing

The Next Wave: Network Publishing

By 2003*:

- Wireless Internet: 350 million
- Households with broadband: 40 million
- 16 billion Web pages
- 90% of sites will be personalized



*estimated

Source: Datamonitor, Ovum, Lyra, IDC, Adobe Market Research

Desktop PC/ Laser Printer | Web | **Internet, Ecommerce More Bandwidth, New Devices**

Source: Datamonitor, Ovum, Lyra, IDC, Adobe Market Research

Today's Challenges: Parallel Workflows

- Redundant content creation
- Device-specific workflows
- Inefficient collaboration
- Generic content delivery

HTML

- **HTML is a “markup language”**
  - “Hypertext Markup Language”

What are markup languages and XML?

- **HTML is a Markup Language**
  - "HyperText Markup Language"
- **XML is NOT a Markup Language (in spite of what the name says)**
  - eXtensible Markup language
- **XML is a standardized way to create Markup Languages**
  - Many, many Markup Languages…

XML for…

- Letters
- Business Cards
- Technical Manuals
- Graphics
- Financial Information
- Protocols
- Metadata

XML Markup Languages

- SVG – XML for Scalable Vector Graphics
- OEB – XML for Electronic Books
- SMIL – XML for Synchronized Multimedia Integration Language
- JDF – XML for Job Definition Format (Print Job Tickets)
- XBRL – XML for Business Reporting (SEC disclosure)
- NewsML – XML for News Stories
- ebXML – XML for Electronic Commerce Transactions (from EDI)
- DocBook – XML for Software Documentation (from SGML)
- XHTML – XML for web browsers

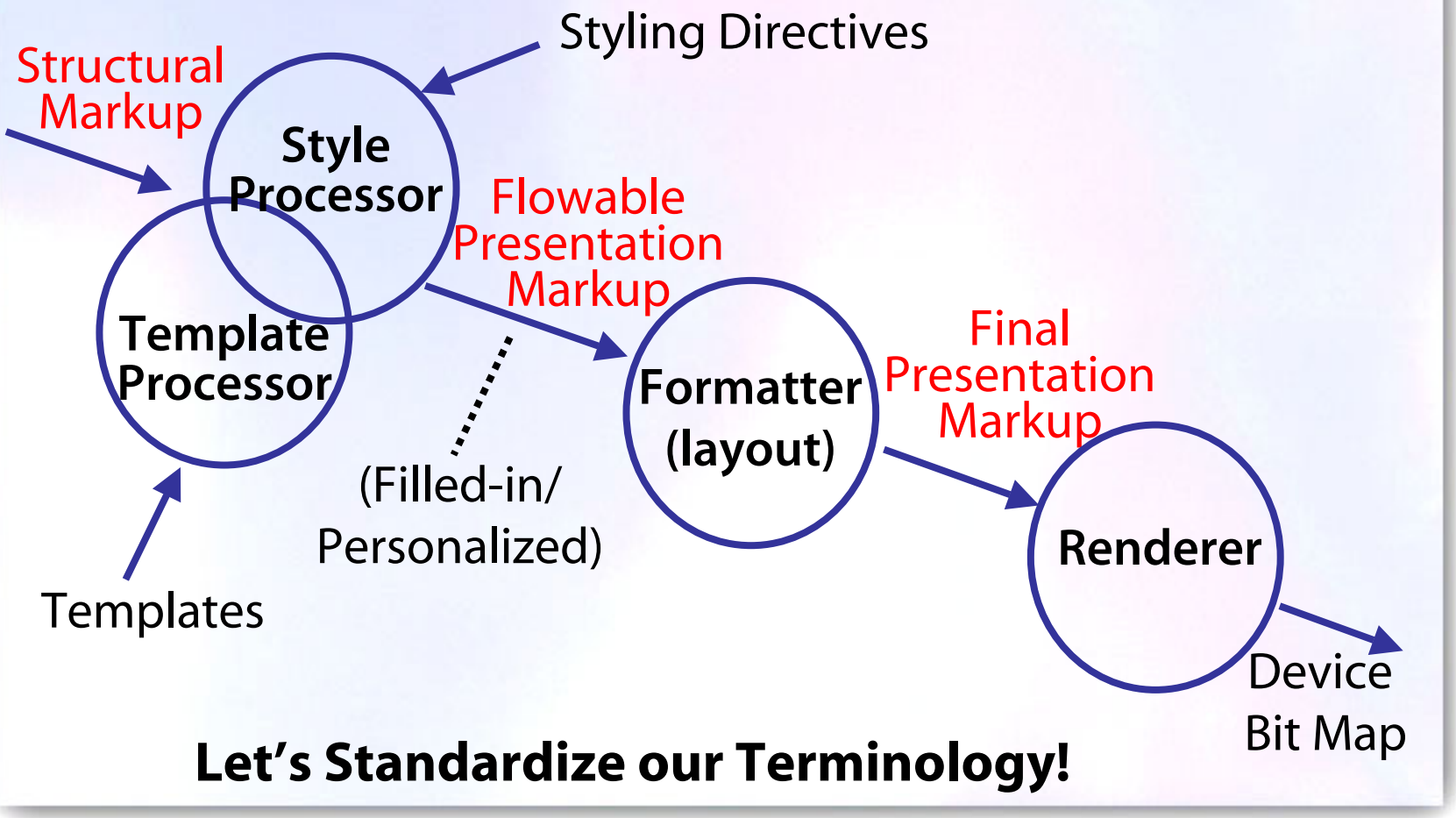
Or, check the "XML Catalog" at http:// www.xml.org

To Understand These Standards and Roles in Network Publishing

We need an Abstract Model

Processing Markup (for Presentation)

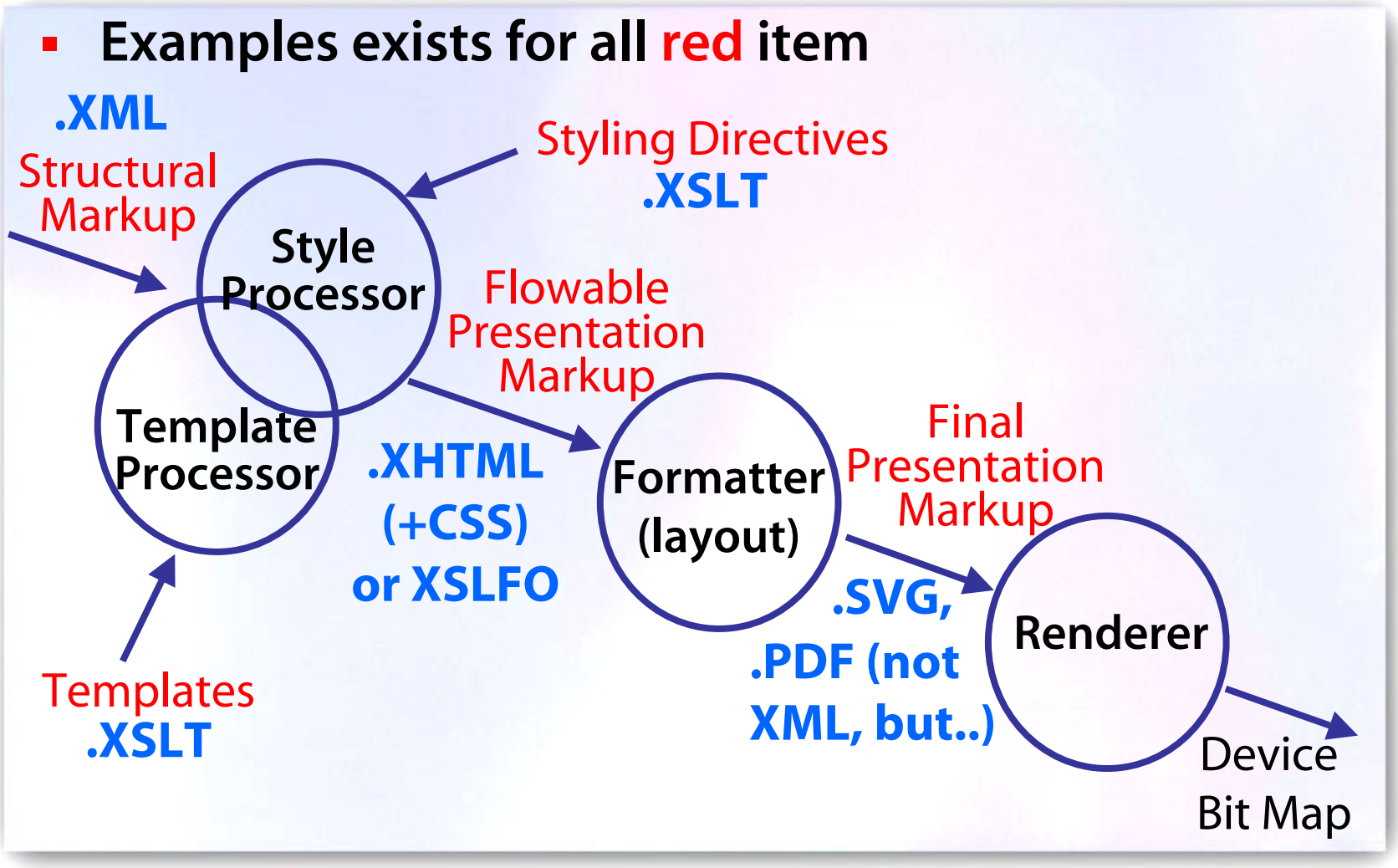
- From Markup to Bit Maps



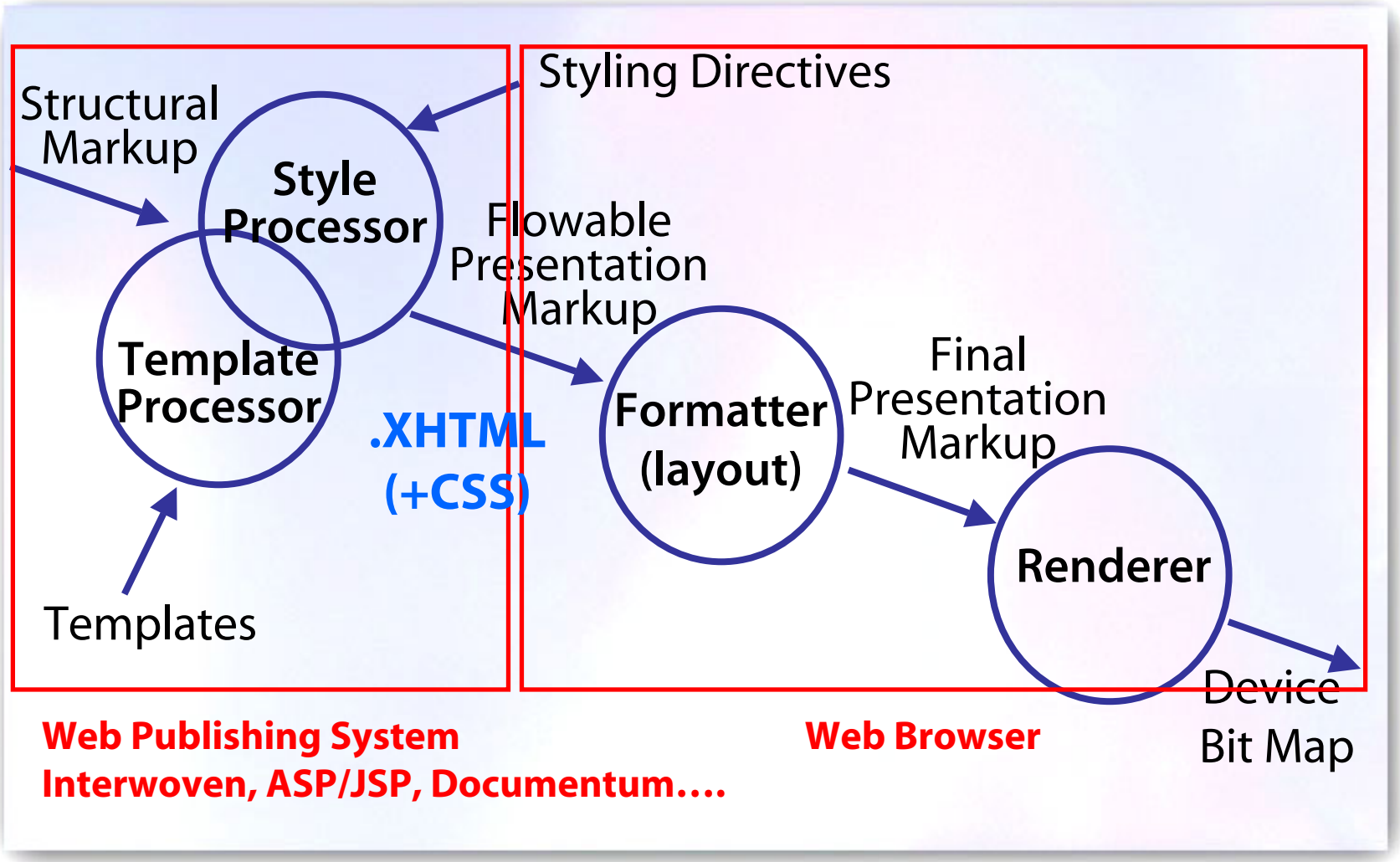
Let's Standardize our Terminology!

XML Markup Languages

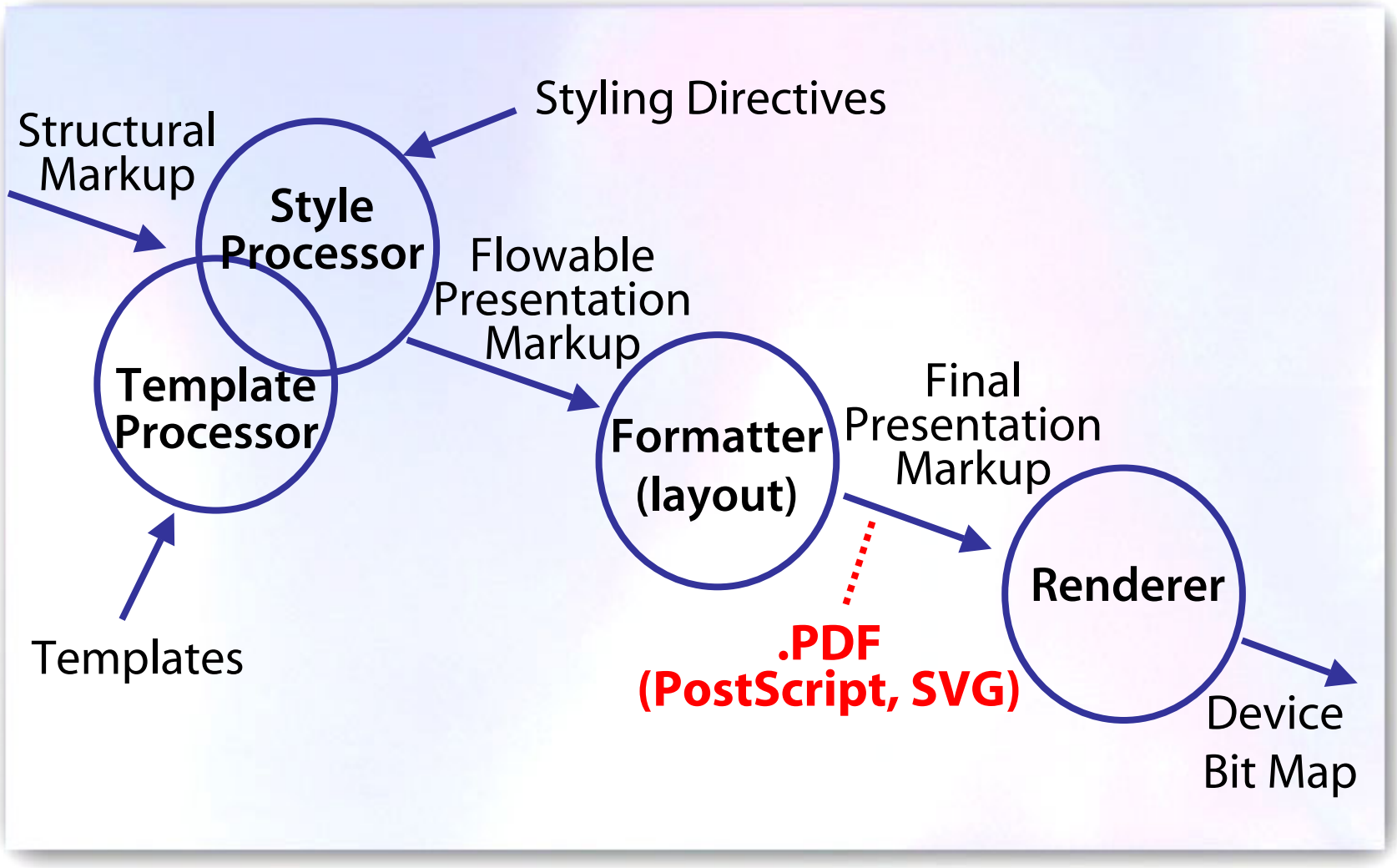
- Examples exists for all red item



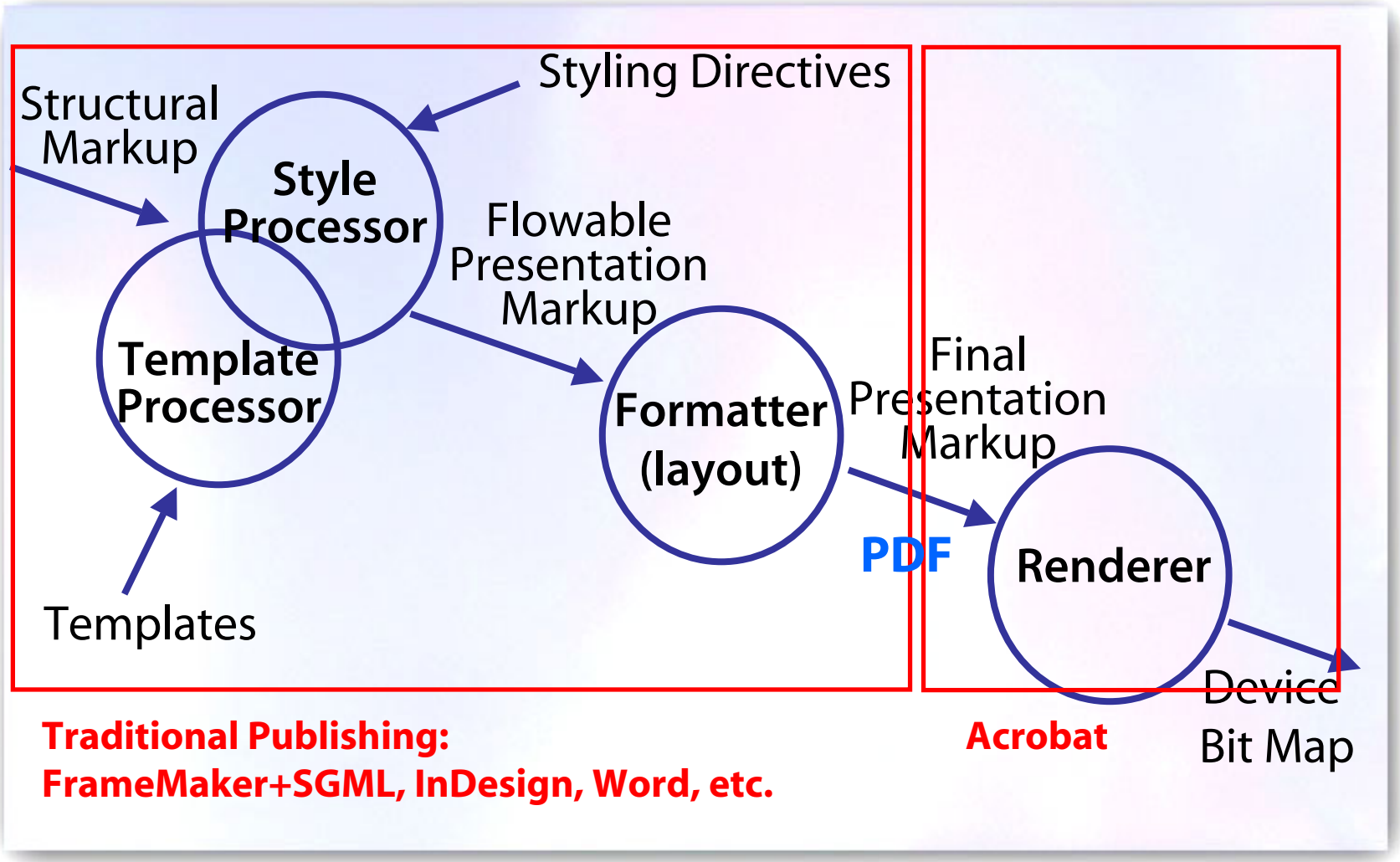
XML Formatting Tools: Web Publishing



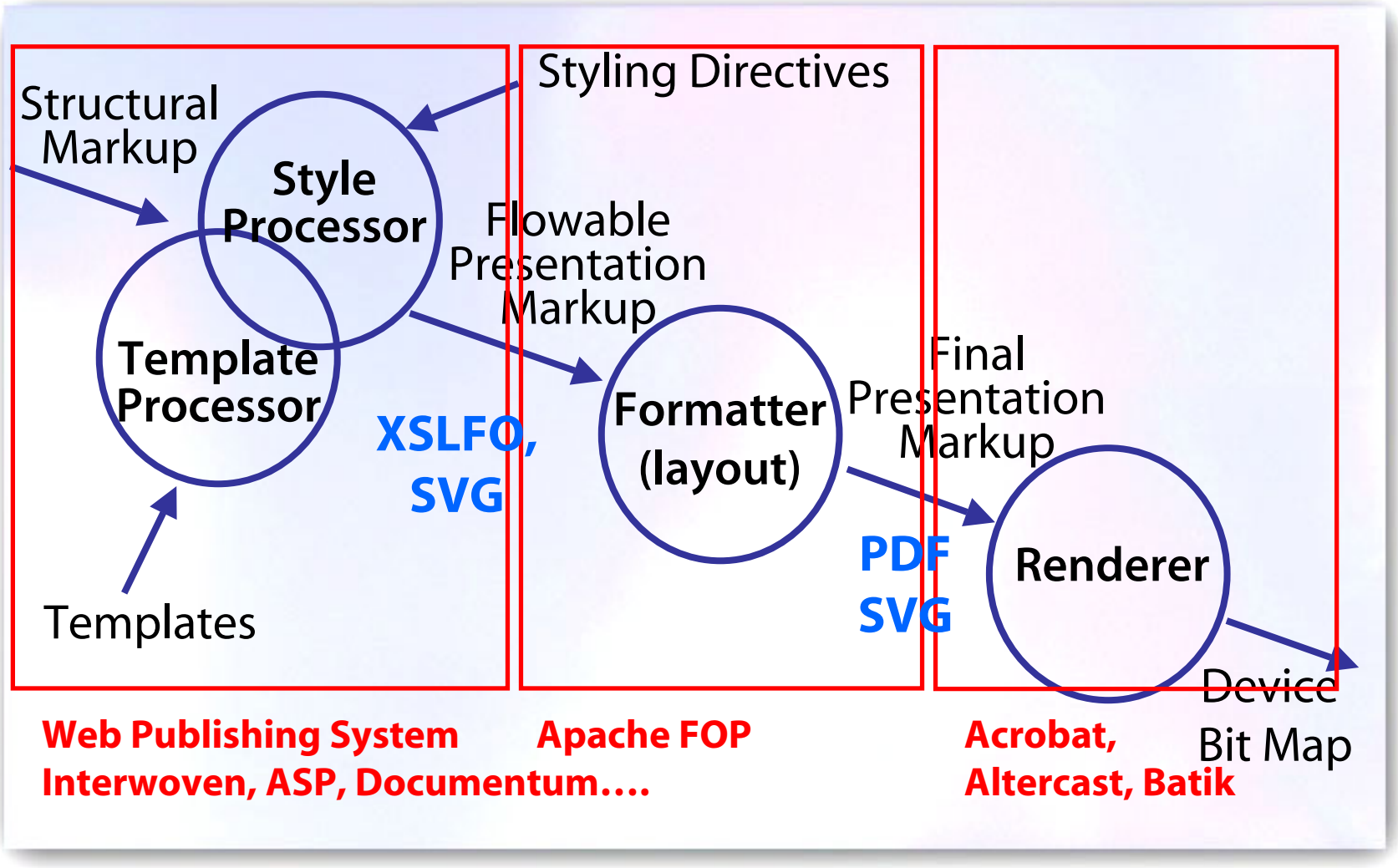
PDF in Processing Workflow



XML Formatting Tools: Traditional Publishing



XML Formatting Tools: New Models - XSL Formatting Object, Server Process



And the most common question is:

"Should I do XML or PDF when publishing?"

Should I Use This?

Sheet Music

Screenplay

Ingredients and Recipe

Wire Service

Tagged Data, Database

Or This?

CD, MP3

Broadway, DVD

Cake

Newspaper

Report, Bill, Book

The Answer is Both!

Sheet Music	CD, MP3
Screenplay	Broadway, DVD
Ingredients and Recipe	Cake
Wire Service	Newspaper
Tagged Data, Database	Report, Bill, Book

Information Distributors Need Both XML and PDF

- **XML is for Structured Information**
  - Describes what elements are and do
  - Main virtue: Flexibility (=adaptability)
- **PDF is for Electronic Page Packages**
  - Describes appearance of typeset page
  - Main virtue: Inflexibility (=stability)

The bad news: there's no magic button

- **XML takes work—defining what you want and tagging what you've got**
- **PDF takes work—defining the design and formatting the pages**
- **The good news: there's real consensus**
  - XML frees data from proprietary codes
  - PDF frees pages from display problems, retains your desired style

XML at Adobe

- **It is being implemented with nearly everything…**
  - InDesign, FrameMaker+SGML, PageMaker
  - GoLive – Dynamic Link
  - SVG – Illustrator, Altercast
  - Photoshop, Illustrator
  - JDF
  - Metadata (XMP), Asset Management, PRISM
  - Standards usage: ICE, Soap, Webdav
  - Acrobat…

Adobe's XML Taxonomy

- XML for data
- XML for metadata
- XML for presentation: display and delivery
- XML for arbitrary content storage

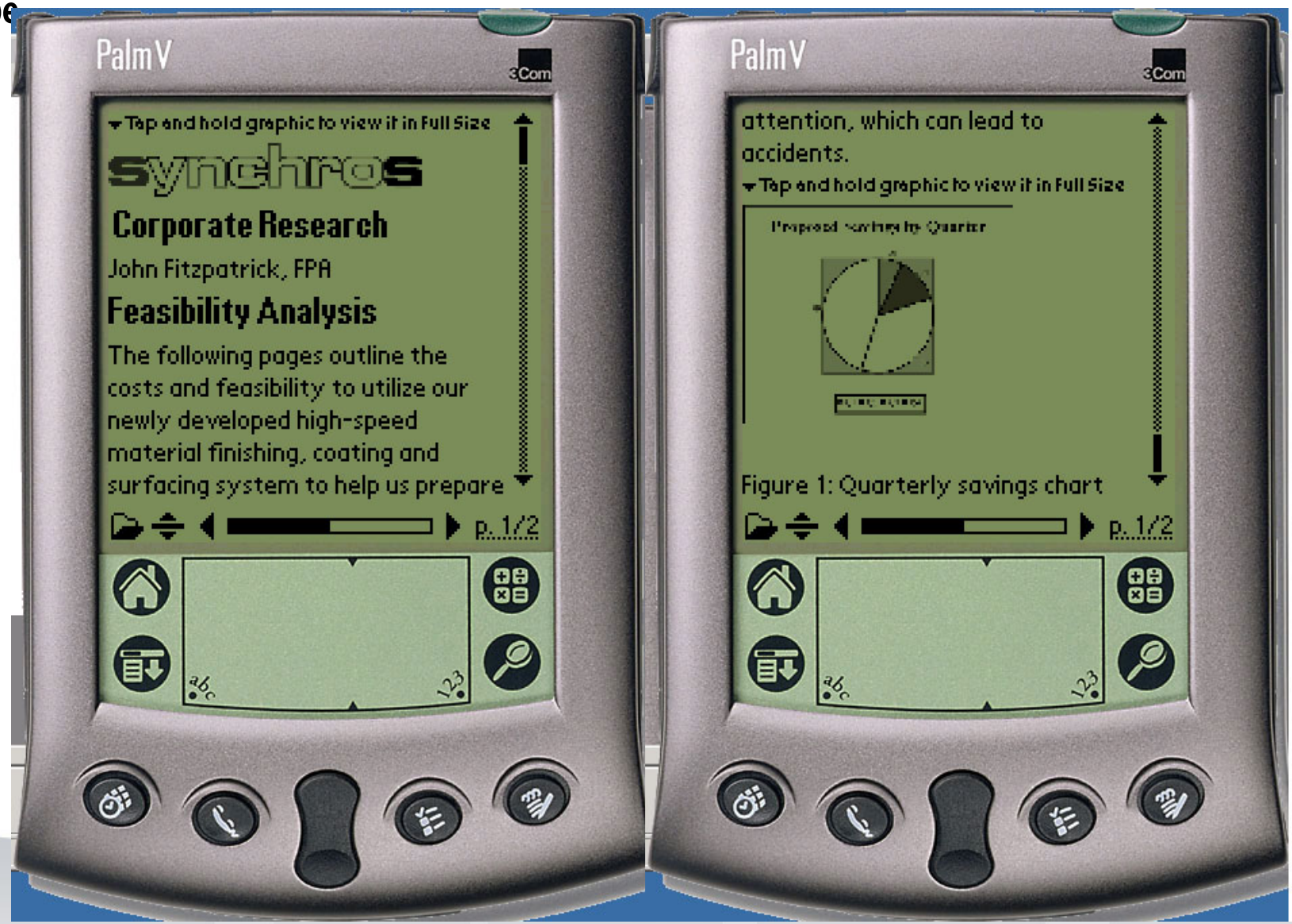
Focus on XML within Acrobat 5.0

- **XML for forms data (XFDF)**
- **XML for embedded data**
- **XML for metadata (XMP, RDF)**
- **XML for content structure in PDF (Tagged PDF)**
  - Structure for reflowable display on desktop PCs, PDAs and other handheld devices, such as eBooks
  - Structure for accessibility
  - Structure for Saveas RTF (XML, HTML in beta)

Demonstration

- XML forms data
- Tagged PDF Creation - Word
- Tagged PDF Usage
- SaveAs XML
- XMP Metadata
- Tagged PDF – Arbitrary XML

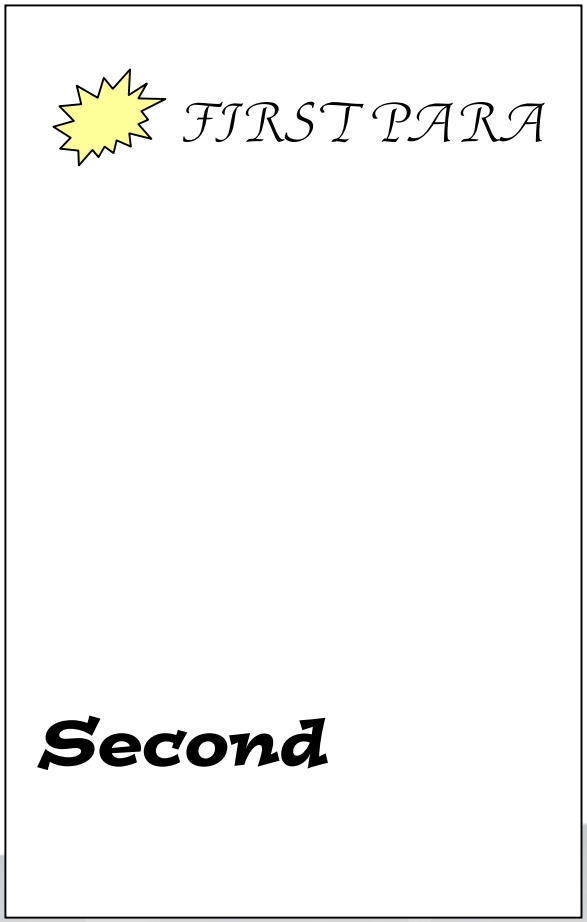
Palm Pilot PDF Demonstration

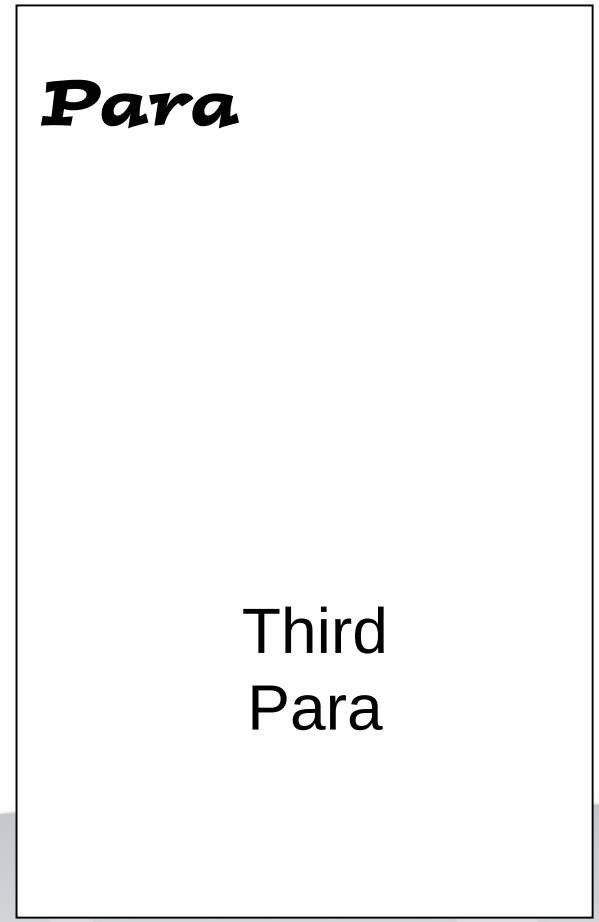


Tagged PDF Visual Example
XML View (illustrative)

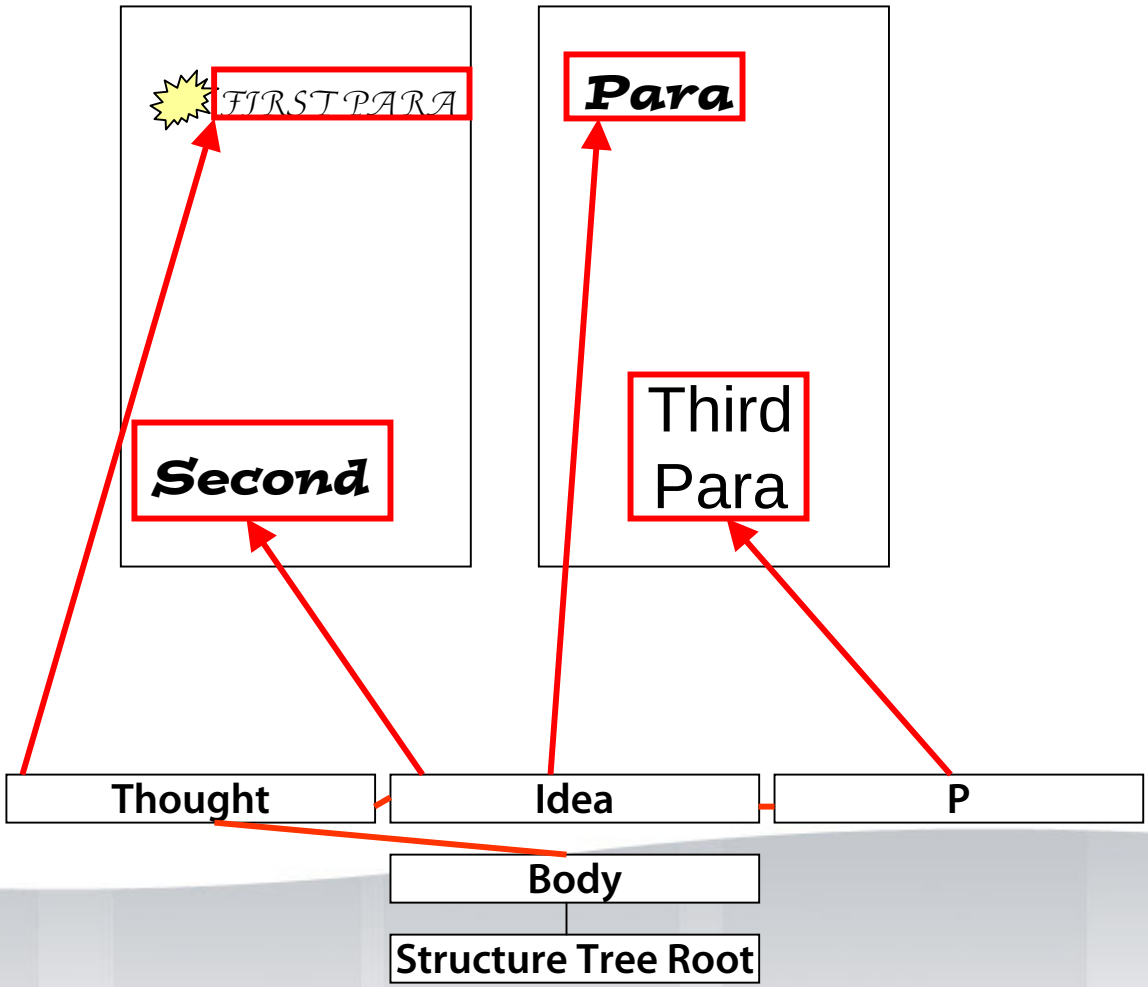
```
<?xml version="1.0" encoding="UTF-8" ?>
<Body>
<Thought>First Para</Thought>
<Idea>Second <?page-break>Para</Idea>
<P>Third Para</P>
</Body>
```

Tagged PDF Visual Example
Page View

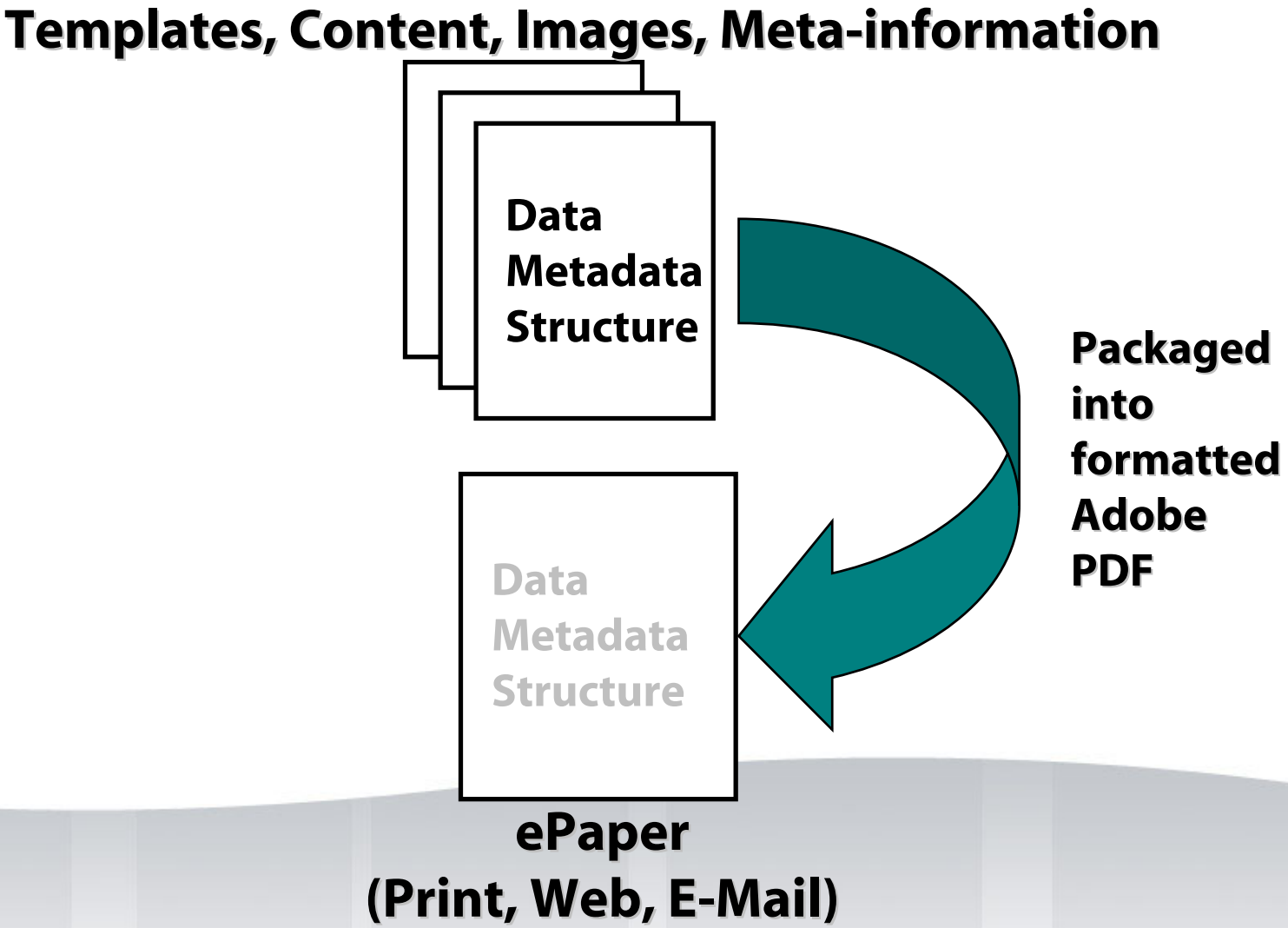




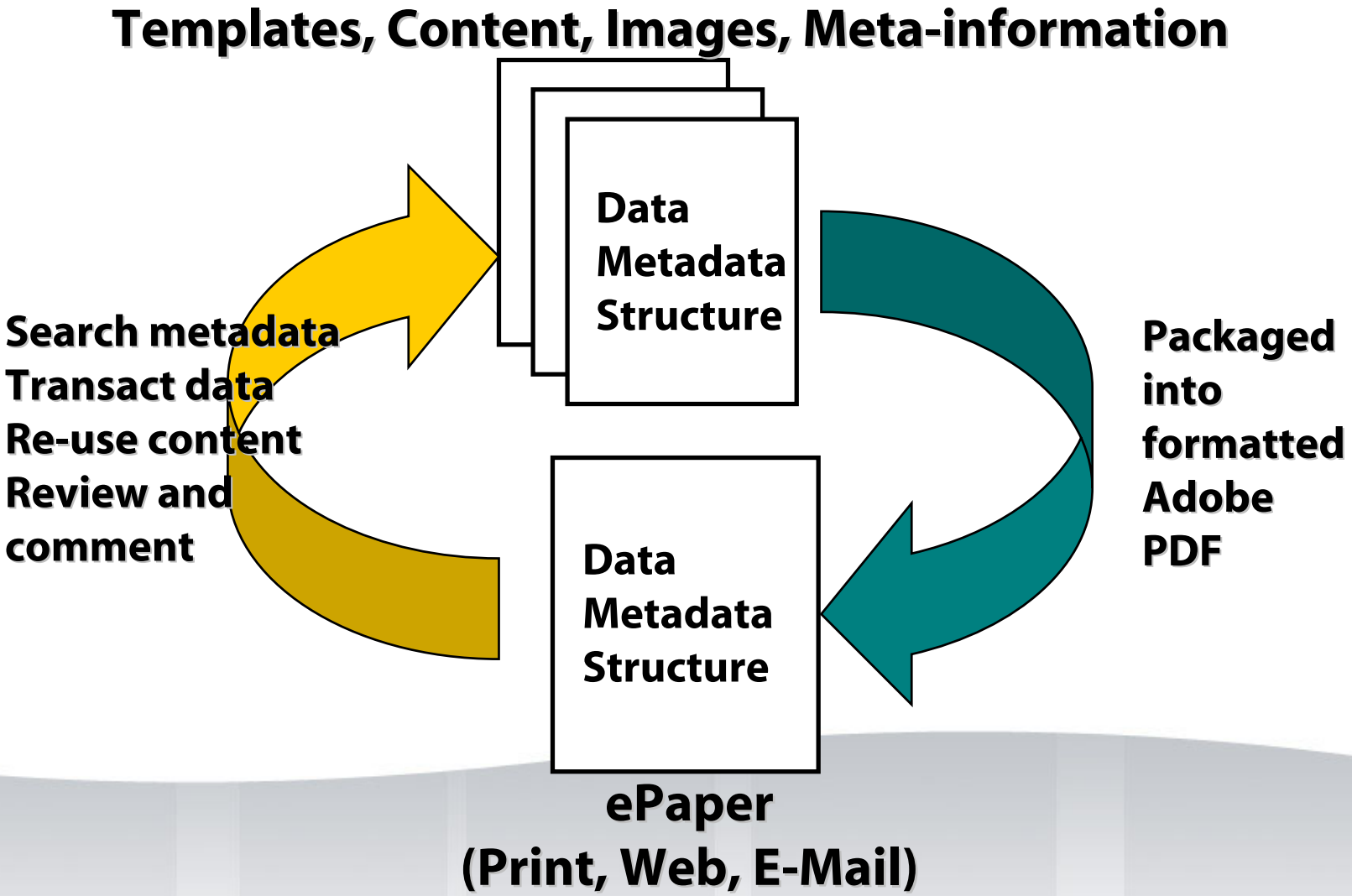
Tagged PDF Visual Example
Page/Structure View



Earlier Versions of Acrobat Enabled Paper to Become ePaper



Acrobat 5.0 Makes Adobe PDF a Full Citizen in eBusiness Processes



PDF in Parallel Workflows

- Page-serving, URL links to page
- Reliable Web delivery (on and offline)
- Reliable Printing
- Access to Impaired
- Handheld Access

Methods to Create Tagged PDF files

- Adobe products –
  - Tags panel/ touchup for manual insertion (rudimentary operations)
  - Acrobat 5/ Office 2000, WebCapture, MakeAccessible plugin
  - Authoring tools - PageMaker
    - Future – InDesign, FrameMaker
- PostScript generators - Distiller/ pdfmark
- Acrobat API (& PDF Libraries)
- Direct PDF generation

XML and PDF Integration Summary

- XML is neutral source format, for driving web, print, wireless
- XML needs formatting and presentation, which Adobe tools provide
- PDF is the best multi-platform, device independent container for formatted representation of XML
- PDF preserves information value and metadata from XML and other sources
- XML and RDF (XMP) give clear integration points between Processes/Repositories and PDF
- XML and PDF, not XML or PDF

Adobe Resources

- **Acrobat 5 SDK**
  - PDF 1.4 Reference Manual
  - pdfmark Reference Manual
  - Acrobat Core API Reference
    - Section 9 - PDSEdit - Creating and Editing Logical Structure
- **Adobe Solutions Network (ASN)**
  - http://partners.adobe.com/asn/developer/acrosdk/acrobat.html
  - http://partners.adobe.com/asn/webseminars (tagged PDF details 11/28, 10 AM PST)

Web Resources

- **Make Accessible plugin (and doc)**
  - http://www.adobe.com/support/downloads/88de.htm
  - http://access.adobe.com/whitepaper.html
- **Saveas Beta plugin on Adobe**
  - http://www.adobe.com/support/downloads/main.html
- **Acrobat Reader for PalmOS/PocketPC**
  - http://www.adobe.com/products/acrobat/readerforpalm.html
- **www.adobe.com for Acrobat**
  - http://www.adobe.com/epaper/main.html
- **XML and PDF: Why We Need Both**
  - http://www.impressions.com/resources_pgs/SGML_pgs/XML_PDF.pdf

Adobe

Q&A

Adobe

everywhere
you look™