

Publisher Conversion Strategy

BookExpo America 2001

**Ken Brooks, President
Publishing Dimensions
kbrooks@pubdimensions.com**

Inputs can be grouped into 3 categories

<u>Input Type</u>	<u>Sub-Type</u>	<u>Example</u>
Text	Straight text	ASCII
	Structured	HTML, XML, OEB
	Styled (formatted)	RTF, Word, WordPerfect, typesetting (3B2, etc.)
	Structured and styled	XML w/CSS, HTML w/CSS, SGML
Electronic Pages	Application files	Quark, InDesign, PageMaker, Frame, Ventura, LaTeX
	PDF	
Fixed Pages	Hardcopy	Books,
	Scan	TIFF, PDF-Image
	Other	Microfilm, printing film, etc.

There are three output categories as well

Format Family

- XML-based

- PDF

- Image

Variant

- DocBook, DocBook Lite ¹
- TEI, TEI Lite ¹
- OEB, RHI Public Implementation
- OEB, netLibrary spec
- HTML4.0
- XHTML
- Gemstar (RB, SB, ReB 1100, 1200)
- MSR (.LIT)
- Palm Digital Media (PML)
- MobiPocket (OEB)
- AER optimized
- Standard PDF
- POD PDF
- PDF-Image
- TIFF (single and multi-page)

¹ Not covered in this presentation

Successful conversion entails mastering numerous combinations

	<u>PDF</u>	<u>OEB/HTML-based eBooks</u>	<u>Image</u>
Paper	✓	✓	✓
Electronic Pages	✓	✓	n/a
Text	✓	✓	n/a

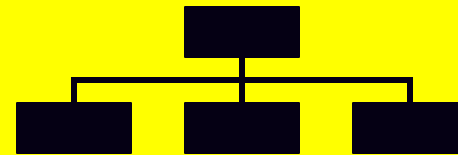
Aside: Conversion requires effective handling of all components of a source document

Content

<p>To be or not to be, that is the question. Whether 'tis nobler ...</p>

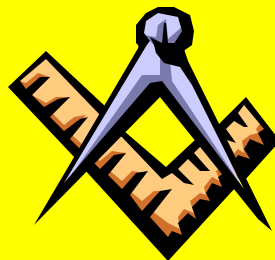
- Tagged “chunks”
- Text, images, tables, equations, etc.

Structure



- DTDs

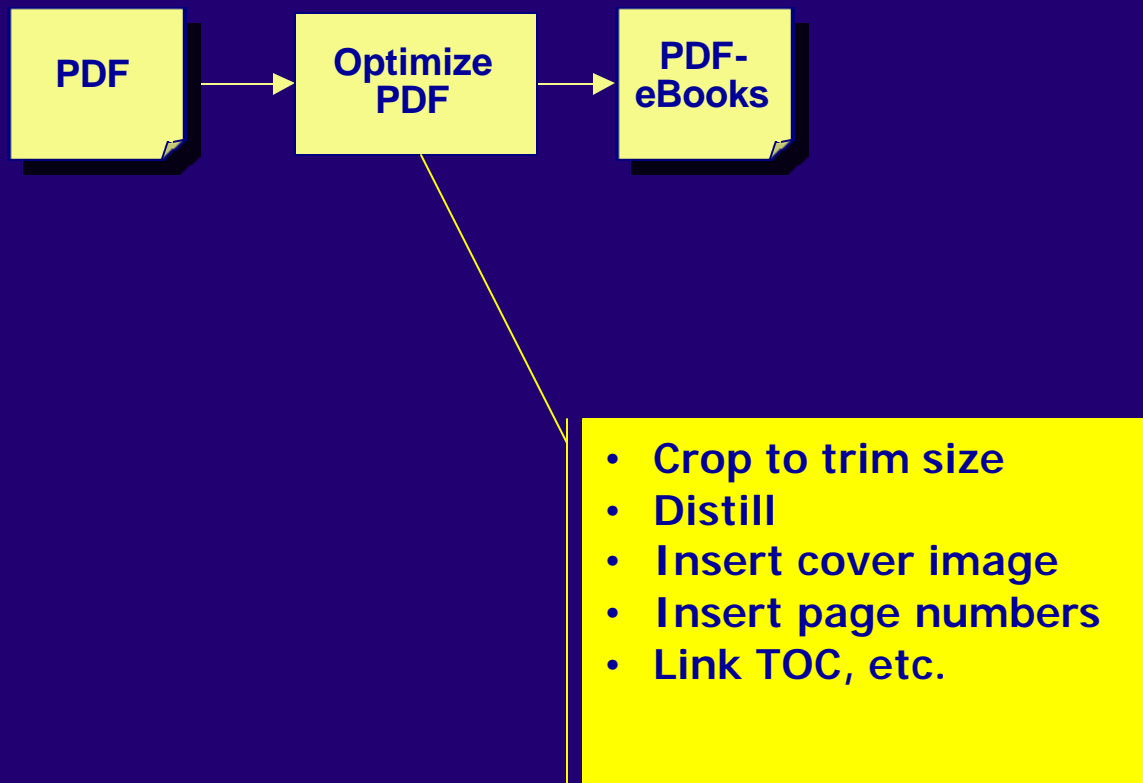
Design



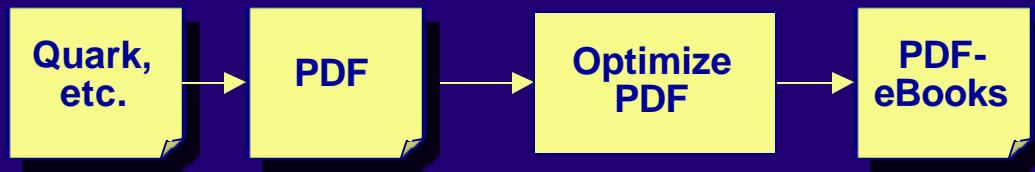
- Master pages
- Styles
- Type specs.



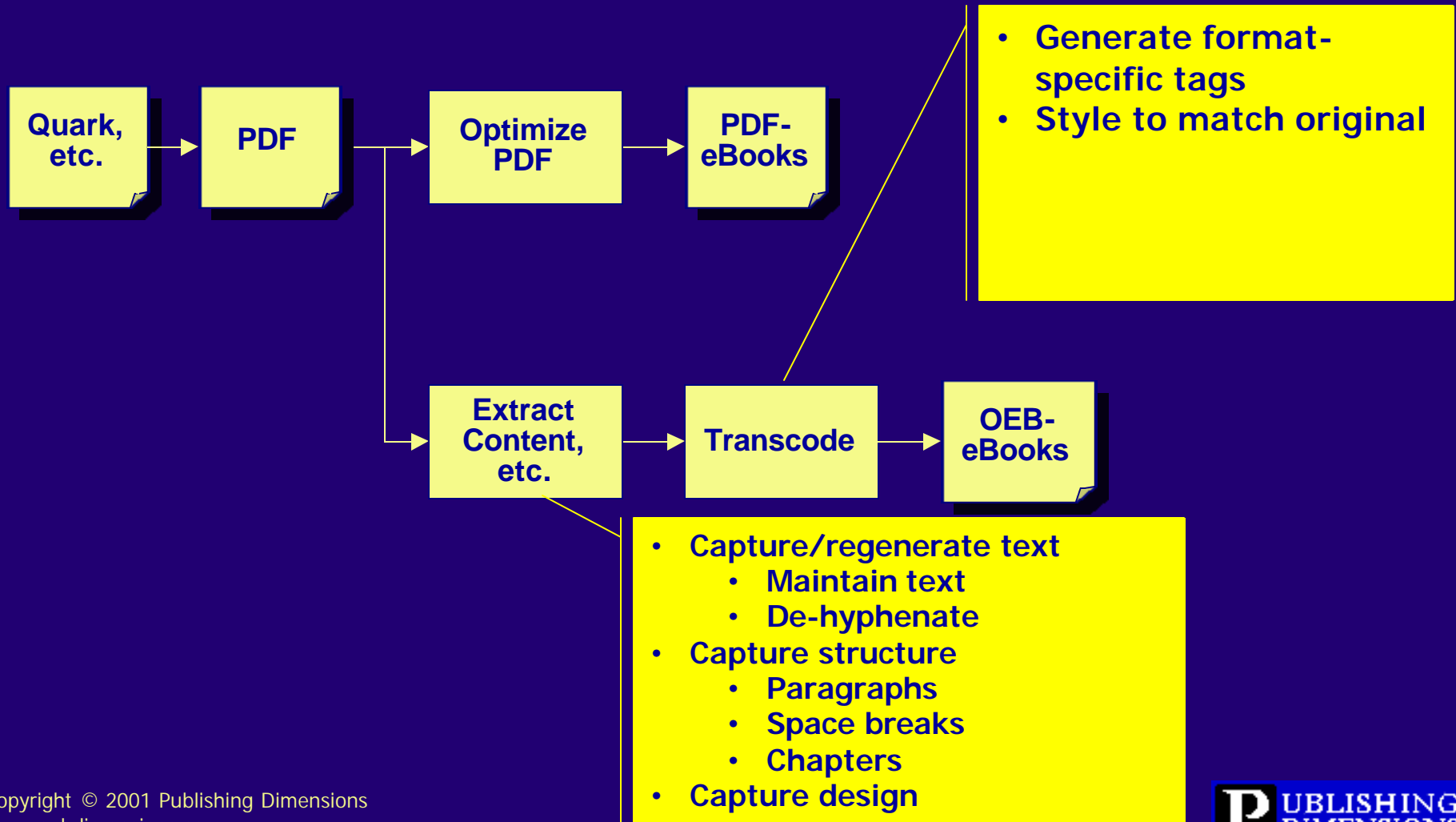
Electronic Pages: Conversion of archived PDFs to PDF eBook formats is straightforward...



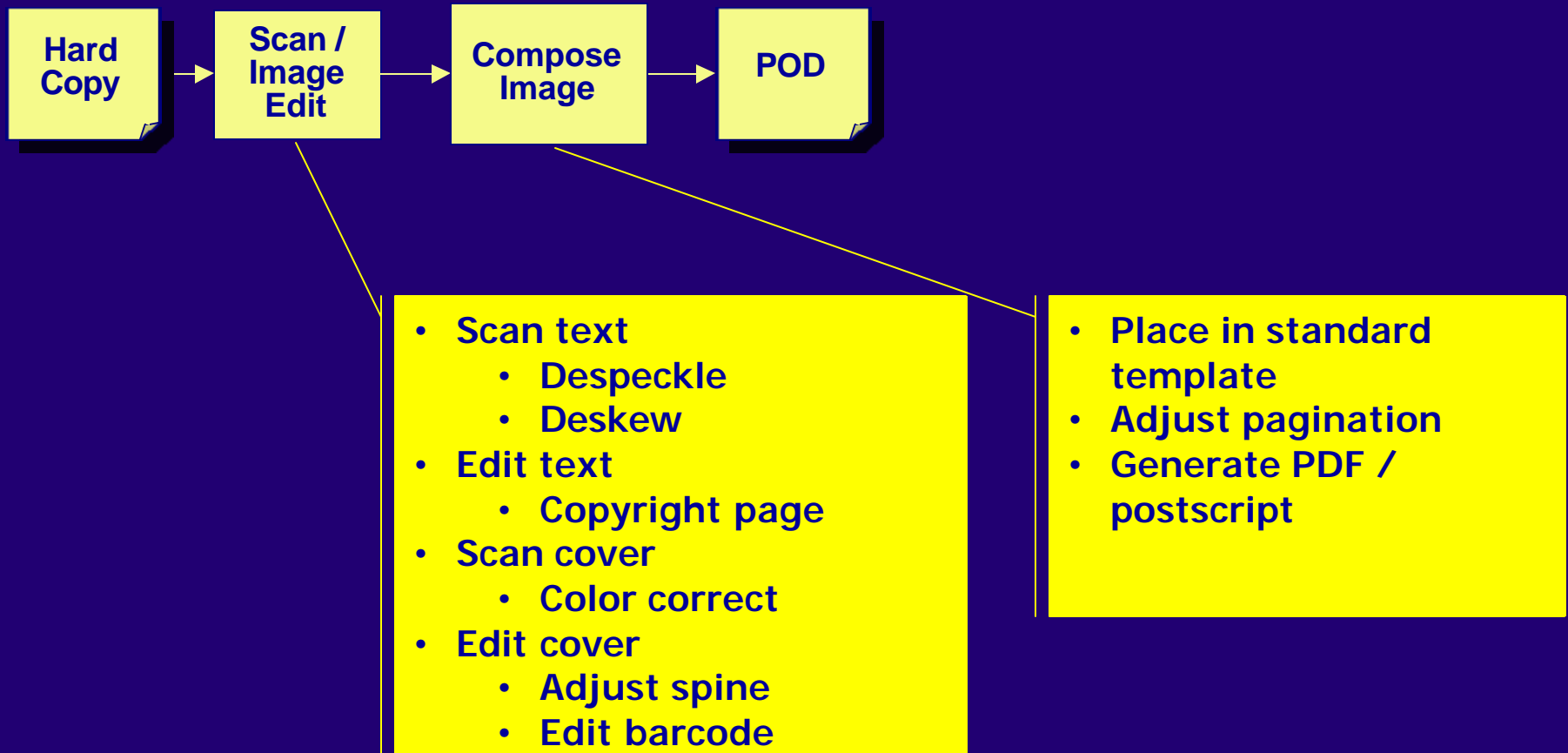
Electronic Pages: ...and from page makeup files to PDF-based eBooks only slightly more complex...



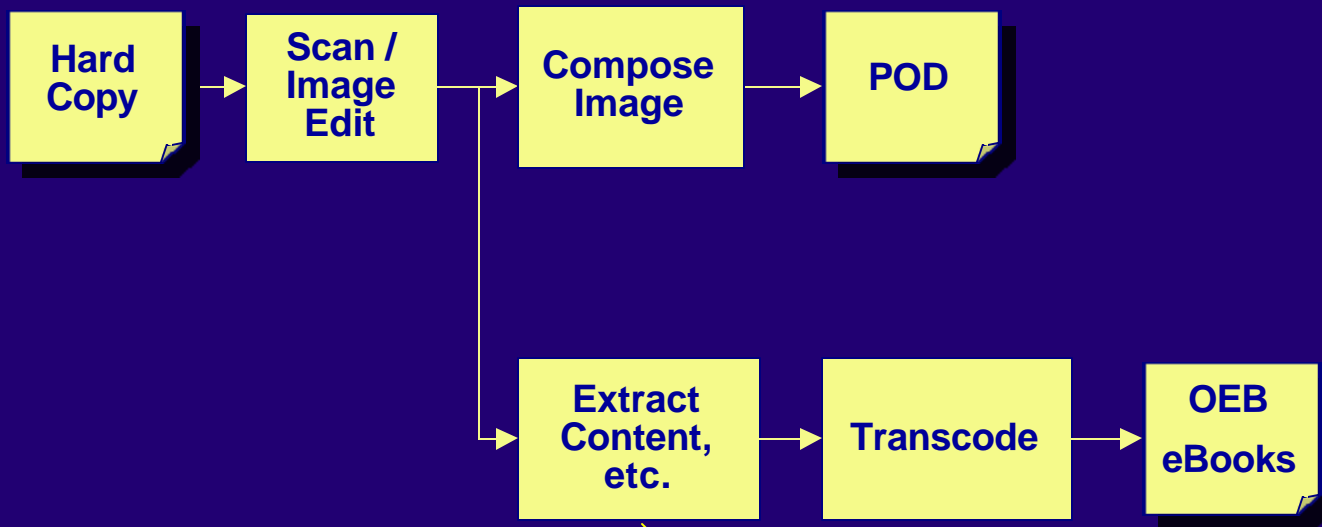
Electronic Pages: ...and to XML-based eBooks also relatively straightforward.



Fixed pages: Hardcopy conversion to POD is also straightforward...

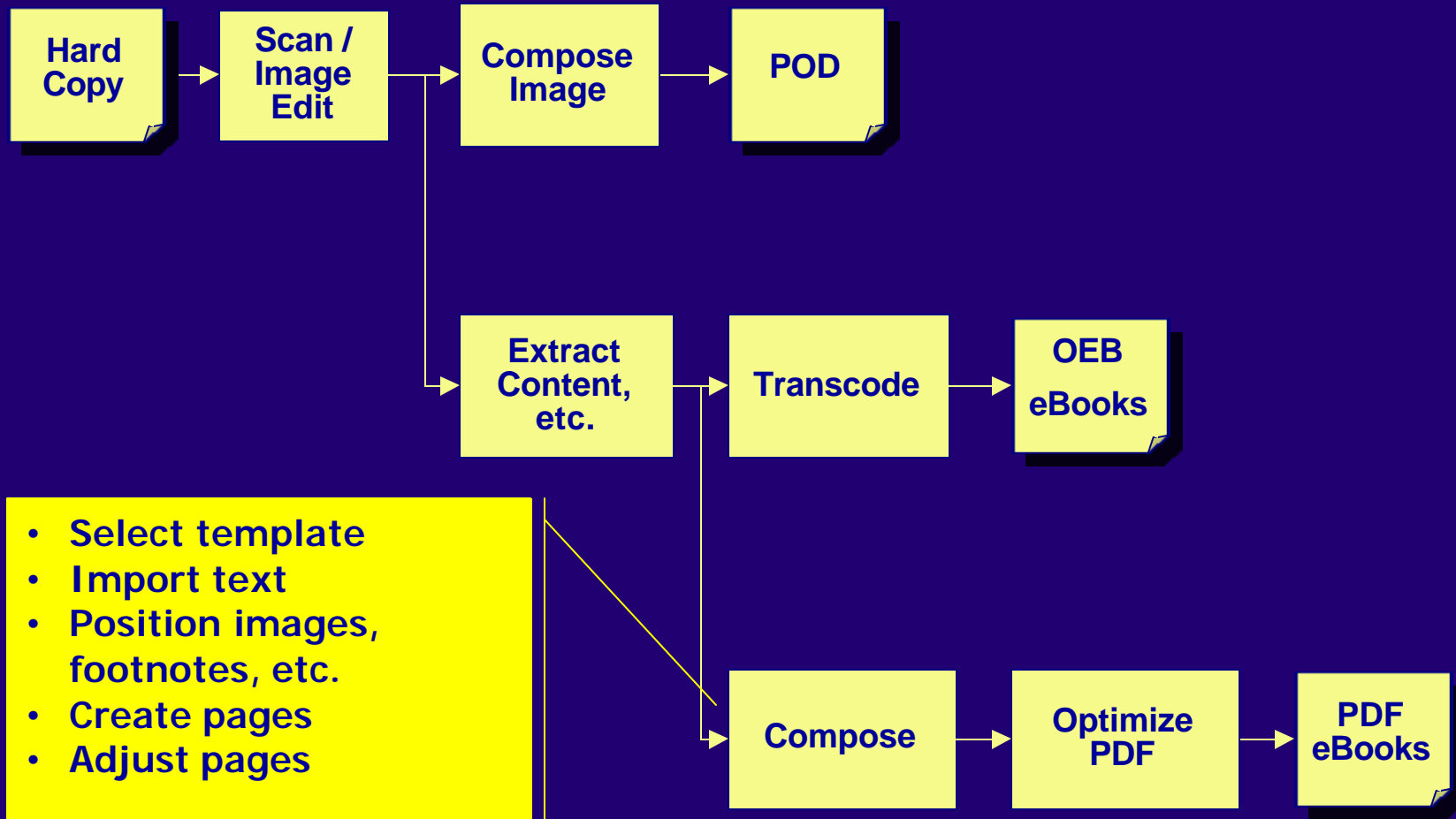


Fixed pages: ...but to XML-based eBooks a bit more involved...

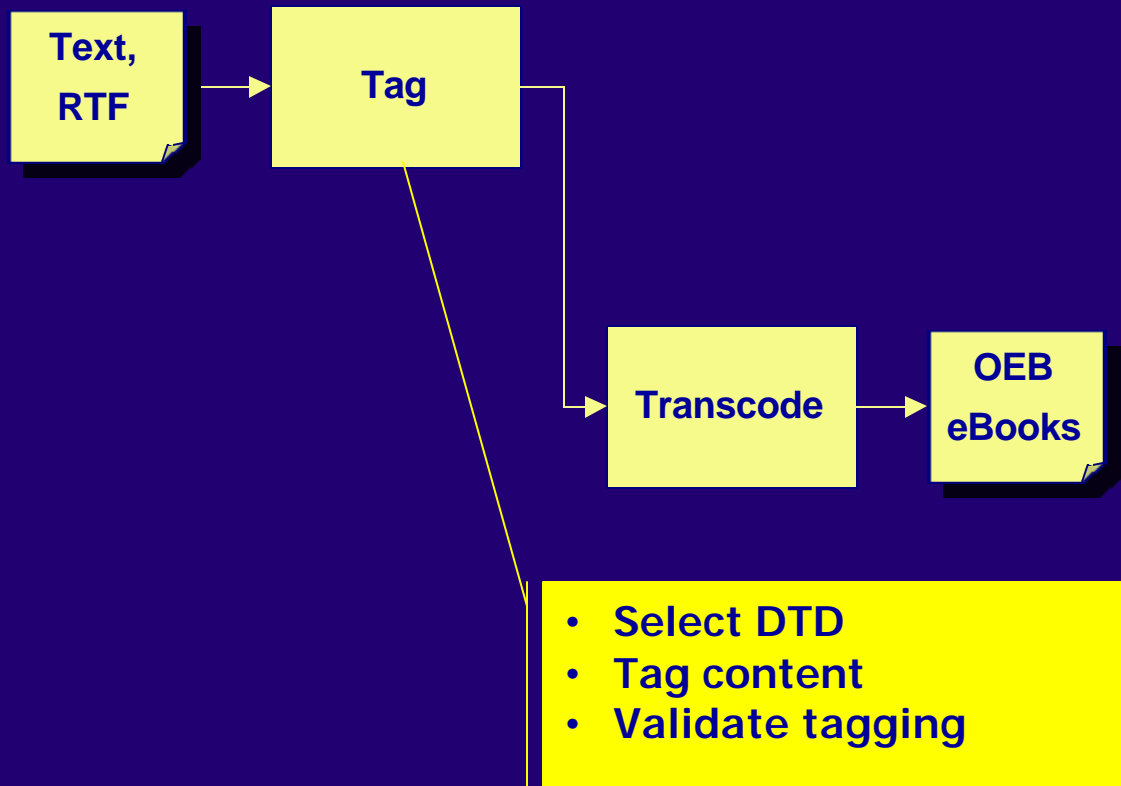


- Extract text, images, etc.
 - Apply OCR
 - Edit (manual, rule-base)
 - (Proofread)
- Extract structure
 - Select DTD
 - Tag and validate
- Capture design

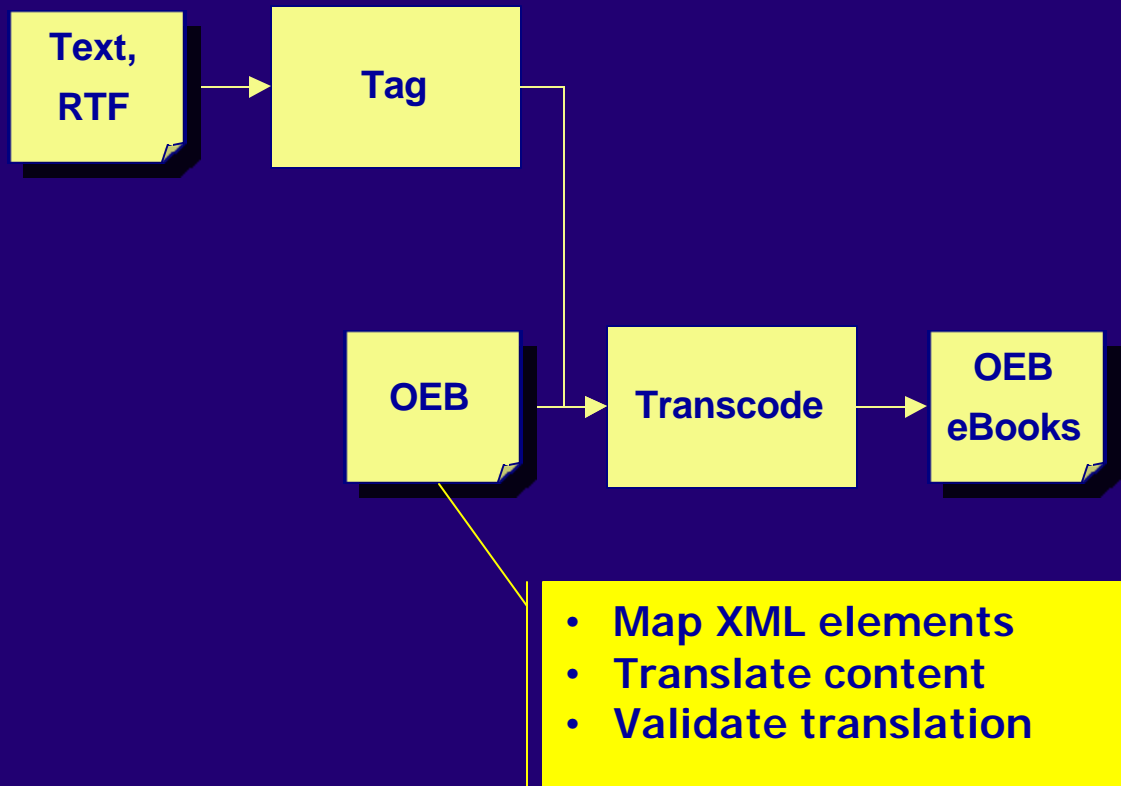
Fixed pages: ...and to PDF-based eBooks, more involved still.



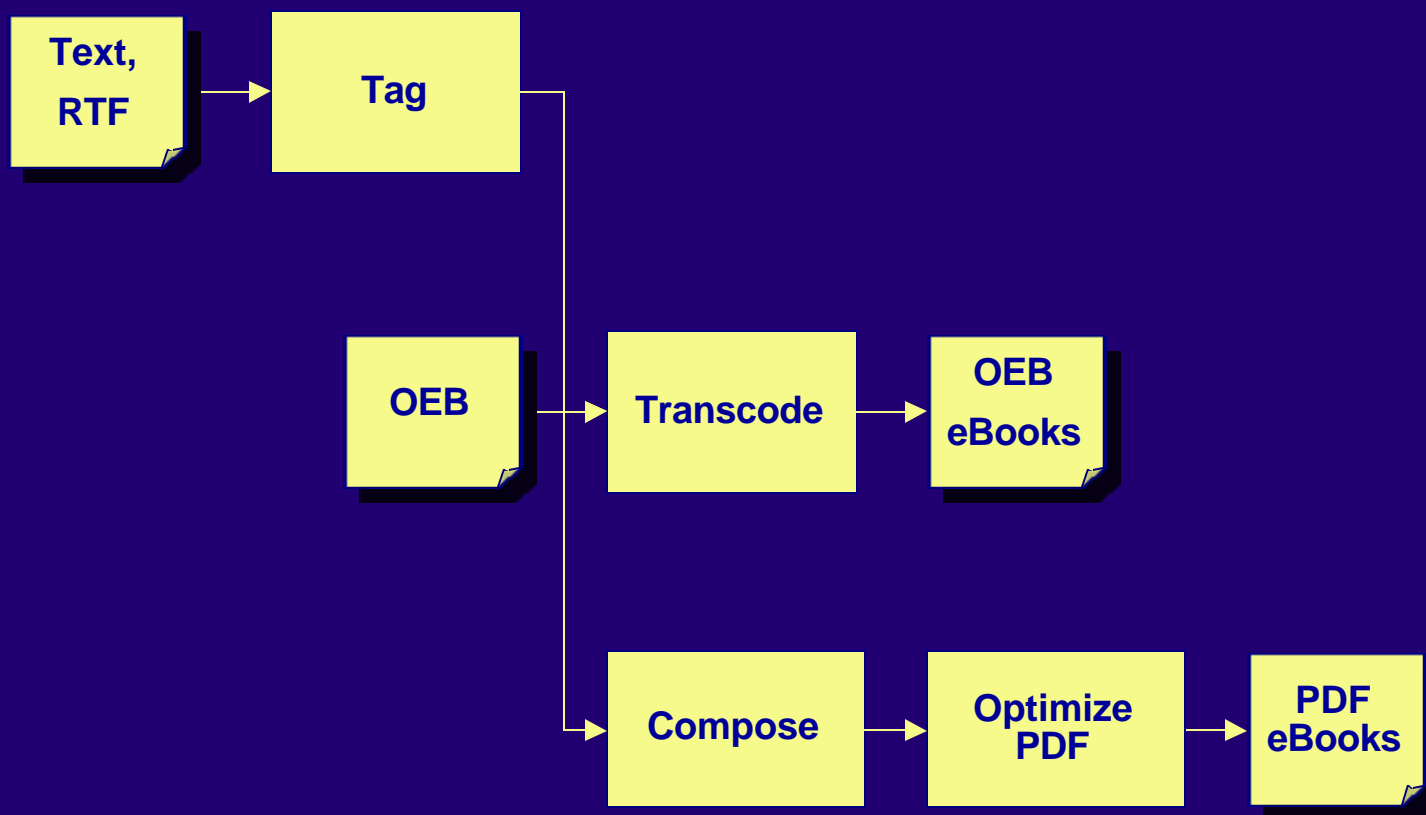
Text: Text input involves very little effort in text capture, and slightly more in structure capture



Text: With OEB input capturing structure is largely a matter of mapping XML elements.



Text: Creation of PDFs requires paging



The priorities

Type of Conversion		List Affected	Breakeven Volume	Increment. Breakeven
Electronic	POD, AER	Frontlist (Backlist)	25	25
	MS, Palm, etc.	Frontlist (Backlist)	37	12
Hardcopy	POD	Backlist	28	28
	MS, Palm, etc.	Backlist	77	54
	AER	Backlist	102	48

Publisher Conversion Strategy

BookExpo America 2001

**Ken Brooks, President
Publishing Dimensions
kbrooks@pubdimensions.com**