

Wrangling the Backlist

XML & PDF to EPUB conversion

...and other thorny issues of making legacy content available digitally

Sriram Panchanathan
SVP, Digital Solutions
Aptara

- **Backlist Economics**
- **Rights to Site – Process Primer**
- **Automation**
- **Content Design and Enhancement**
- **Quality**
- **Decision Factors**
- **Scalable and Cost Effective Management**
- **About Aptara**

- 10-15 million print backlist, ~1 million in e-Book format
- Only 20% of your content will pay off but which 20%?
- The sum of selection is greater than the parts
- Quality conversion and optimal user experience costs \$\$\$
- Cost of e-book = cost of rights + cost of conversion + cost of distribution
- UX, quality, selection cost \$\$\$ and also bring in \$\$\$

Acquire

- Rights Acquisition
- Source Acquisition
- Project Management

Convert

- Design
- Process Setup
- Automated Conversion
- Manual Conversion
- Automated and Manual Quality Control
- Product Delivery

Distribute

- E-Tailer Account Setup
- Metadata Management
- Content and Metadata Upload
- Payments Receipt
- Promotions
- Monitoring and Ongoing Management

**Every step adds to cost.
Rights & source acquisition is often a bigger
bottleneck than conversion**

Technology Process

Page Image to XML

- Scanning
- OCR
- Page zone auto recognition
- Page pattern recognition

XML to eBook

- XSLT
- Pattern matching heuristics
- AI applications

- No automated solution can guarantee 99.9X% quality
- Software does not handle unstructured data well enough (yet)
- “Autonomation” should complement Automation
- Cannot get acceptable quality at \$20 a book of conversion (mostly)

Key Questions

- Should baseline UX match print experience?
- Should I enhance and what should I enhance?
- What assets are available?
- How should I layout the book?
- How to render complex elements and multimedia?
- How to style? What should CSS look like?

Effective Content Design =

Multimedia Web UX Design + Editorial/Composition + HTML/CSS/Javascript

- Quality can be measured through % accuracy
- Types of defects
- Quality factors
- Recommendations:
 - every title should have a baseline level of quality
 - automation at all stages eliminates defects
 - only manual touch can guarantee highest level
 - tradeoff Quality/Cost/Revenue while investing

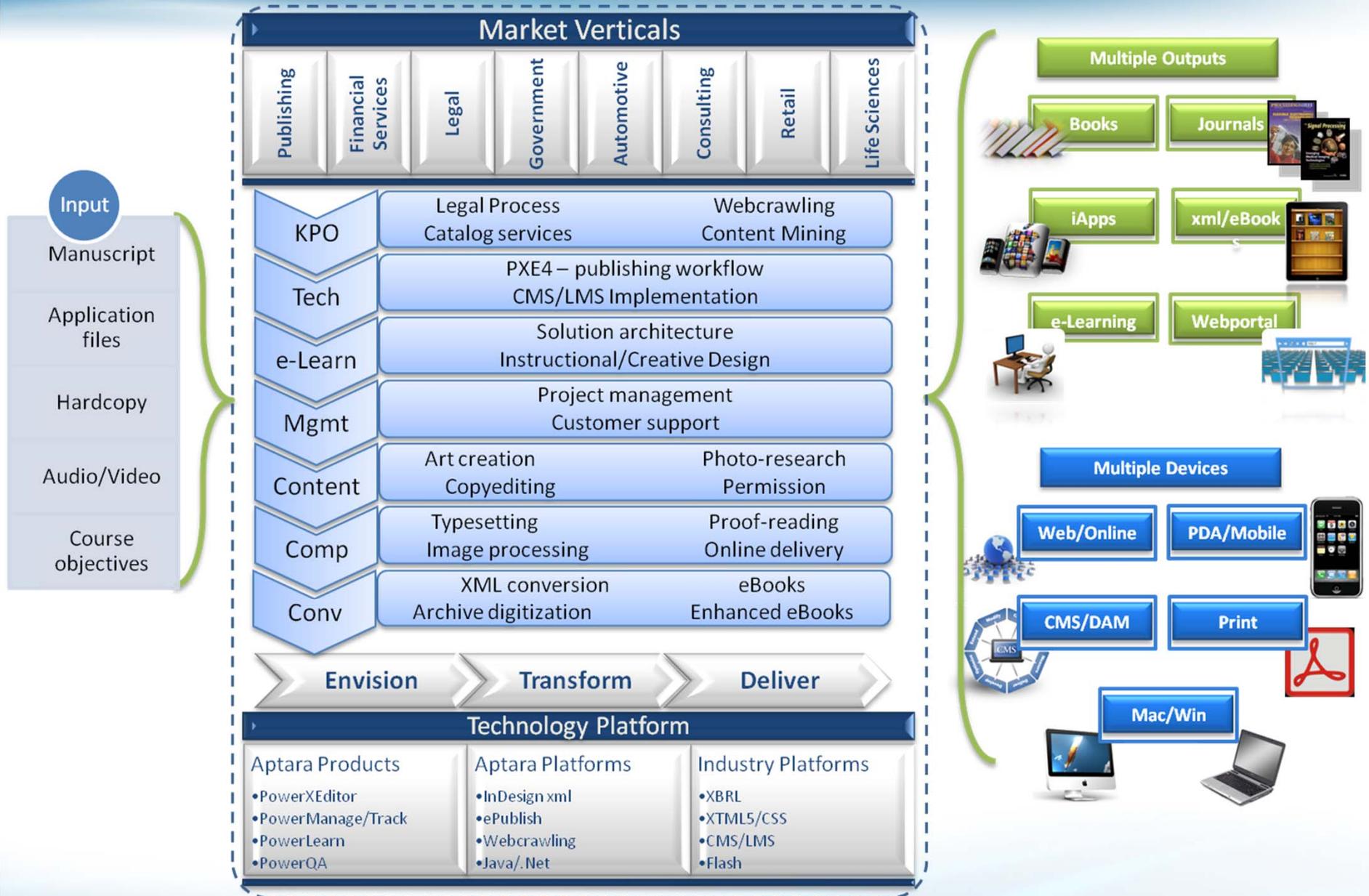
- Source quality
- Source complexity
- Rights availability
- Output quality expectation
- Output design complexity
- Selection value
- Revenue potential = $f(\text{print sales, page views, gut feel})$
- Business Value = $f(\text{selection value, revenue potential})$

Business Value	High	No Brainer!!	Bucketize and attack easy buckets
	Low	Judiciously drop feature+quality bar!!	Save it for another day!!
		Low	High
		Complexity of Conversion	

Start with a budget, 'bucketize" and prioritize your backlist and attack the easy buckets first!

- Production strategy should be in-line with your marketing strategy
- Convert more than what qualifies to spread costs
- You need a baseline level of quality for all content
- Your ability to scale is only as good as your virtual supply chain
- Choose a technology partner who offers:
 - flexibility
 - scale
 - end to end capabilities
 - and can deliver volumes without affecting quality

Aptara – End to End Content Transformation Solutions



Thank You

Sriram Panchanathan
sriram@aptaracorp.com

302-465-3922

703-352-0001

BACKUP

Accuracy % (A)

$$A = 1 - \frac{X}{T}$$

where

A : Accuracy %

X : number of defects

T : total number of items

Example :

X = 325, T =
12,500

A = 1 - (325/12500)
= 1 - 0.026 = 97.4%

Where

% Accuracy Means	Defects/Page
99%	20 Defects/Page
99.9%	2 Defects/Page
99.99%	1 Defect/5 Pages
99.995%	1 Defect/Page
99.998%	1 Defect/25 Pages



99% vs 99.9% Accuracy

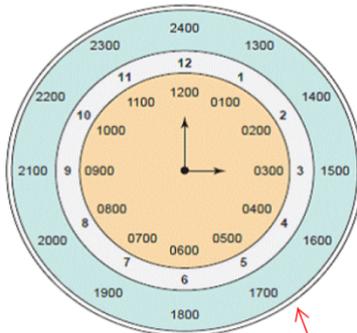
For most of their long marriage, Bill Piper's career as a chemical engineer with DuPont had anchored them in Delaware. Rosie had enjoyed their time in the South. But with the realization that she alone could not manage his needs, she decided on the Boston area, where their two daughters had settled. Mary Beth, the rebellious one no longer went to church; but her husband, Peter, attended Mass with Rosie.

apostrophe missing
hyphen missing
rebellious should be "rebellious."

short-term currency trading, and converted during World War II as a depository for captured Axis loot. The fund held \$200 million earmarked for the reconstruction of Europe. It delivered millions into the bank accounts of wealthy American citizens. Americans, who then sent the money to the U.S. Its created by the CIA. Donors were instructed to place a special code on their income tax forms alongside their charitable donation. The millions were delivered to Italian politicians and priests of Catholic Action, a political arm of the Vatican.¹¹

Between the CIA and US relief projects, America poured \$350 million into Italy during the twelve months before the April 18, 1948 election. The Vatican Bank, which was founded in 1942 under Pius XII to consolidate the Holy See's worst of World War II to the Christian Democrats. Pius "provided 100 million lire bank," writes John Cornwell, "sum of money apparently raised from the sale of U.S. war materiel and earmarked for the Vatican to spend on anti-Communist activities." The 1948 parliamentary elections were the first under Italy's postwar constitution since the fall of Fascism.

Conventional time	24-hour time
4:00 A.M.	0400
5:00 A.M.	0500
6:00 A.M.	0600
7:00 A.M.	0700
8:00 A.M.	0800
9:00 A.M.	0900
10:00 A.M.	1000
11:00 A.M.	1100
12:00 P.M. (noon)	1200
1:00 P.M.	1300
2:00 P.M.	1400
3:00 P.M.	1500
4:00 P.M.	1600
5:00 P.M.	1700
6:00 P.M.	1800
7:00 P.M.	1900
8:00 P.M.	2000
9:00 P.M.	2100
10:00 P.M.	2200
11:00 P.M.	2300
12:00 A.M. (midnight)	2400



emdash should be hyphen
remove hyphen
link missing
remove space
"a" missing
remove space
space is missing

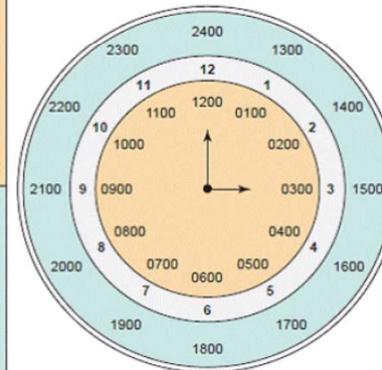
Like meat balls, it's too small a THING to have its own department unlike Rudeness and Selfishness which are located down various aisles of the store known as Marriage. should be lower case "thing"

For most of their long marriage, Bill Piper's career as a chemical engineer with DuPont had anchored them in Delaware. Rosie had enjoyed their time in the South. But with the realization that she alone could not manage his needs, she decided on the Boston area, where their two daughters had settled. Mary Beth, the rebellious one, no longer went to church; but her husband, Peter, attended Mass with Rosie.

short-term currency trading, and converted during World War II as a depository for captured Axis loot. The fund held \$200 million earmarked for the reconstruction of Europe. It delivered millions into the bank accounts of wealthy American citizens, many of them Italian Americans, who then sent the money to newly formed political fronts created by the CIA. Donors were instructed to place a special code on their income tax forms alongside their "charitable donation." The millions were delivered to Italian politicians and priests of Catholic Action, a political arm of the Vatican."

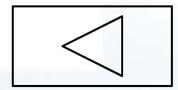
Between the CIA and U.S. relief projects, America poured \$350 million into Italy during the twelve months before the April 18, 1948, election. The Vatican Bank, which was founded in 1942 under Pius XII to consolidate the Holy See's finances during the worst of World War II, became a conduit for funds to the Christian Democrats. Pius "provided 100 million lire [\$185,000] from his personal bank," writes John Cornwell, "a sum of money apparently raised from the sale of U.S. war materiel and earmarked for the Vatican to spend on anti-Communist activities." The 1948 parliamentary elections were the first under Italy's postwar constitution since the fall of Fascism.

Conventional time	24-hour time
1:00 A.M.	0100
2:00 A.M.	0200
3:00 A.M.	0300
4:00 A.M.	0400
5:00 A.M.	0500
6:00 A.M.	0600
7:00 A.M.	0700
8:00 A.M.	0800
9:00 A.M.	0900
10:00 A.M.	1000
11:00 A.M.	1100
12:00 P.M. (noon)	1200
1:00 P.M.	1300
2:00 P.M.	1400
3:00 P.M.	1500
4:00 P.M.	1600
5:00 P.M.	1700
6:00 P.M.	1800
7:00 P.M.	1900
8:00 P.M.	2000
9:00 P.M.	2100
10:00 P.M.	2200
11:00 P.M.	2300
12:00 A.M. (midnight)	2400

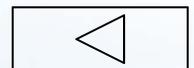


Like meatballs, it's too small a thing to have its own department unlike Rudeness and Selfishness which are located down various aisles of the store known as Marriage.

to push me down to the police station or just out to the parking lot,



Front/Back Matter	<ul style="list-style-type: none">•Cover / Thumbnail•Copyright / TOC / Title / Author name•Naming Configuration
Content	<ul style="list-style-type: none">•Character / Typo•Spacing
Layout	<ul style="list-style-type: none">• Spacing / Formatting• Alignment / Line break•Lists / Note / Dialogue / Footnote / Index
Punctuation	<ul style="list-style-type: none">•Quote•Hyphen•Mdash / Ndash
Bookmark	<ul style="list-style-type: none">•Wrong / Missing•Typo•Link
Special Instructions	<ul style="list-style-type: none">•Book specific
Emphasis	<ul style="list-style-type: none">•Bold / Italic / Font / Font Size / SmallCaps
Images	<ul style="list-style-type: none">•Placement•Clarity
Linking	<ul style="list-style-type: none">•Unwanted•Missing / Wrong



- Source Quality
- Source Complexity
- Output layout complexity
- Multimedia enhancement
- Conversion process effectiveness
- Proofing and Checking effectiveness (automated and manual)
- Time spent/title and hence \$\$\$

