

Word Input to XML workflows

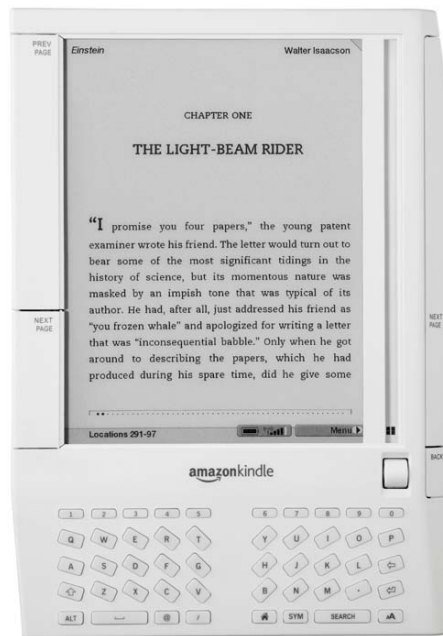
Presented by
Bruce D. Rosenblum
CEO
Inera Incorporated



X-Window XML Publishing Seminar, 6 September 2011

Copyright © 2011 Inera Incorporated. All Rights Reserved

Transformative Technologies...



...Demand New Product Features

- ◆ Automatic reflowable text
 - ◆ Richly hyperlinked
 - ◆ Dynamically updated
 - ◆ Accessible for visually impaired
-
- ◆ Reading a standard PDF on a small screen is not good enough!



New Products Require XML

- ◆ XML drives
 - Composition and PDF
 - HTML
 - ePub
 - Content reuse
- ◆ Single-source production
 - Most efficient
 - Fewest errors



Today's Publications

- ◆ Most journals do XML today
- ◆ Books really *are* different from journals
 - Some core production issues are identical
 - But most book publishers do not use XML
- ◆ Yuck!
 - (We know you've been trying to avoid it for years)
 - (Sorry... you can't hide any longer)



XML Is Not Easy

- ◆ XML (done right) requires
 - New workflow
 - New tools
 - New training
- ◆ XML is a software issue



Where In The Workflow?

- ◆ You can introduce XML at:
 - Authoring
 - Before editing
 - Before composition
 - Post-publication

- ◆ Each point has pros and cons



The Original XML Dream

- ◆ Authors create XML manuscripts
- ◆ Editors edit XML manuscripts
- ◆ XML single-source publication
 - Print
 - eBooks
 - Derivative products



The Author Reality

- ◆ Authors use Microsoft Word, Word Perfect, LaTeX
- ◆ Authors use and “abuse” Word
 - Smart authors “create” solutions
 - E.g. underlined plus (±) for plus-minus (\pm)
- ◆ Documents have the “three uglies”
 - Special characters
 - Tables
 - Math



The Author Reality

◆ Most Authors

- Do not think structure
- Do not like production tasks

◆ Outside Authors

- Brilliant subject matter experts
- Hard to train and support
- Even harder to control
- Can't get IT to install XML editing tools



Post-Publication XML

- ◆ Author submits Word manuscript
- ◆ Edited in Word
- ◆ Typeset (InDesign/Quark)
- ◆ Proof and typeset corrections
- ◆ Publish print and PDF
- ◆ Create XML/ePub from PDF



Post-Publication XML Issues

◆ Advantages

- No workflow changes

◆ Disadvantages

- Quality of XML unchecked
- Extra production time and cost
- Errors discovered in XML creation
- XML sometimes not as richly tagged
- It's not an integrated workflow



Post-Publication ePub Issues

- ◆ ePub created from PDF usually lacks
 - Rich metadata of XML file (e.g. DOI)
 - Internal hyperlinks to footnotes, references, etc.
 - Section 508 accessibility compliance (float positions, table scope attributes, etc.)
- ◆ Broken hyperlinks
 - Especially extracted from multi-column reference lists



XML First Workflow

- ◆ Accept Word manuscript from author
- ◆ Convert manuscript to XML
- ◆ Edit XML manuscript
- ◆ Typeset XML
- ◆ Proof and typeset corrections in XML
- ◆ Create final PDF, ePub, etc.



Advantages and Disadvantages

◆ Advantages

- Only one file conversion
- File is continually validated to DTD

◆ Disadvantages

- Requires XML editing software for all editors
- Training is expensive
- Freelance editors not practical
- Editors work amidst XML tags
 - or XML editing customization is expensive



XML “Middle” Workflow

- ◆ Accept Word manuscript from author
- ◆ Clean up manuscript and style paragraphs
- ◆ Edit in Microsoft Word
- ◆ Convert Word to XML
- ◆ Typeset from XML
- ◆ Proof and typeset corrections
- ◆ Create final PDF, ePub, etc.



Advantages and Disadvantages

◆ Advantages

- Editors work in Microsoft Word
- Lower training costs
- Freelance editors are practical
- Structure enforced prior to final pages

◆ Disadvantages

- Requires running application in-house for XML



Good News – XML Made Easy!

- ◆ Standardized tools make XML-middle workflow available to publishers of all sizes
 - eXtyles – Integrated suite of editorial and XML tools for Microsoft Word
 - Automated XML-driven page layout systems
 - Workflow integration and content management systems
- ◆ All customizable to meet your exact specifications



How eXtyles Helps

◆ eXtyles

- Cleans up Word idiosyncrasies
- Special tools for rapid/accurate style application
- Automates Word technical edit
- Automates markup of semantics
- Easy one-button XML conversion

◆ 90% of conversion is prep

- Other tools don't help with prep



Isn't DOCX XML?

- ◆ Yes and No
- ◆ DOCX captures format in XML
 - RTF with an XML syntax
- ◆ DOCX does not capture semantics
- ◆ Word does not support standard DTDs
 - JATS (NLM), DocBook, TEI, DITA
- ◆ Microsoft's NLM attempt failed



Round-Trip Myth

- ◆ Word → XML → Word?
- ◆ Could be done, but
 - Requires bi-directional conversion tools
 - Where to keep the attributes in Word?
- ◆ XML-middle workflow allows
 - Word master; all editing in Word
 - XML creation at any time
 - Word file returned to author for future editions



Word to XML Summary

- ◆ Document preparation is key
- ◆ Strong tools available
- ◆ Pragmatic workflow required

- ◆ XML can work for you!



Questions?



Bruce Rosenblum
Inera Incorporated
+1 (617) 932-1932
brosenblum@inera.com
www.inera.com

