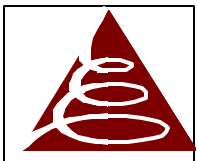


XML and the Production Process

Presented by
Bruce D. Rosenblum
CEO
Inera Incorporated

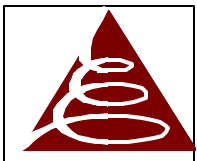
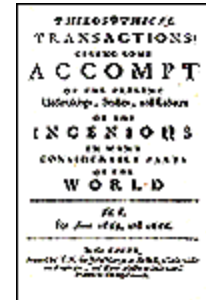
SSP Technology Blitz, 18 November 2004



Copyright © 2004 Inera Incorporated. All Rights Reserved

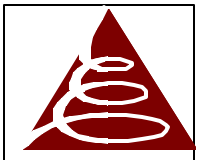
A Little History

- ◆ Gutenberg
- ◆ Oldenburg
- ◆ Linotype
- ◆ Photon
- ◆ PostScript



Last Fifteen Years...

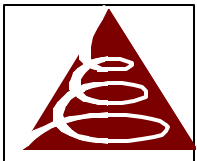
- ◆ More change than the last 500
- ◆ Starting points are different
 - 1989: Paper
 - 2004: Electronic files
- ◆ Ending points are different
 - 1989: Print
 - 2004: Print, PDF, CD-ROM, XML, HTML



New Technologies

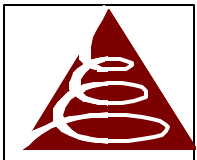
◆ New Opportunities

- Present more dynamic content
- Create new products
- Require new workflows
- Fundamentally transform publishing



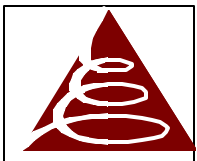
How Radical?

- ◆ A profound revolution in publishing



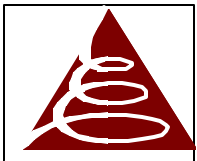
How Radical?

- ◆ A profound revolution in publishing
- ◆ The reign of the Jacobins...



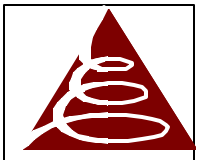
New Challenges

- ◆ Choices?
 - Change



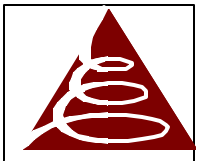
New Challenges

- ◆ Choices?
 - Change
 - Or die...



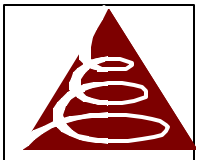
XML Is Not Easy

- ◆ XML requires
 - New workflow
 - New tools
 - New training
- ◆ XML is a software issue



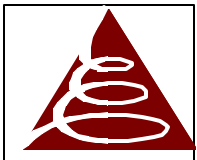
What XML Is and Does

- ◆ XML is a meta language
- ◆ XML drives workflow
- ◆ XML drives the business processes
- ◆ XML drives new products
- ◆ XML drives new knowledge



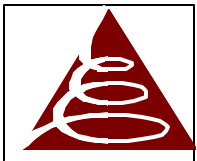
The XML Dream

- ◆ Authors submit XML manuscripts
- ◆ Editors edit XML manuscripts
- ◆ XML single-source publication
 - Print
 - Web
 - Derivative products



The Electronic Reality

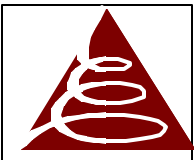
- ◆ Authors submit
 - Microsoft Word
 - Word Perfect
 - LaTeX



The Electronic Reality

- ◆ Authors submit
 - Microsoft Word
 - Word Perfect
 - LaTeX

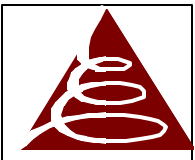
- ◆ La Perfect Word



The Electronic Reality

- ◆ Authors submit
 - Microsoft Word
 - Word Perfect
 - LaTeX

- ◆ La Perfect Word — NOT!



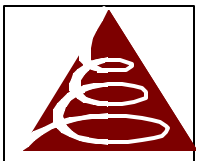
The Author Reality

◆ Most Authors

- Do not think structure
- Do not work linear
- Do not like production tasks

◆ Outside Authors

- Wonderful subject matter experts
- Hard to control
- Hard to train and support



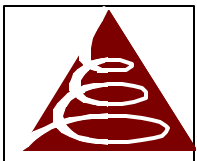
Workflow in Transition

◆ Old World Order

- Paper Manuscript
- Printed Journal

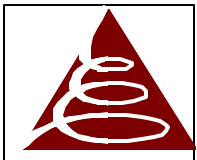
◆ New World Order

- Electronic Manuscript
- Printed Article
- PDF Article
- XML Article



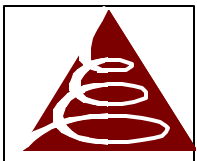
Case Study: Capital City Press

- ◆ Printer of scholarly journals
- ◆ Full service provider
- ◆ Produced full text SGML since 1996
- ◆ Clients include:
 - Elsevier Science
 - Blackwell Science



Original Paper Workflow

- ◆ Submit and edit on paper
- ◆ Keyboard for typesetting
- ◆ Proof
- ◆ Typeset corrections
- ◆ Print



Electronic Workflow, v 1.0

- ◆ Submit electronic or paper
- ◆ Convert to "coded" file
- ◆ Edit coded file
- ◆ Typeset from coded file
- ◆ Re-key tables and math
- ◆ Proof and typeset corrections
- ◆ Print and create PDF
- ◆ Create SGML



Author's File in Word

Nuclear γ -Tubulin during Acentriolar Plant Mitosis

Pavla Binarová^{a*}, Věra Cenklová^b, Bettina Hause^c, Elena Kubátová^a, Martin Lysák^b, Jaroslav Doležel^b, László Bögre^d, and Pavel Dráber^e

^aInstitute of Microbiology, Academy of Sciences of the Czech Republic, Videňská 1083, 142 20 Prague 4, Czech Republic.

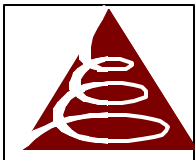
^bInstitute of Experimental Botany, Academy of Sciences of the Czech Republic, Sokolovská 6, 772 00, Olomouc, Czech Republic.

^cInstitute of Plant Biochemistry, P.O.Box 110432, D-06018 Halle, Germany.

^dSchool of Biological Sciences, Royal Holloway, University of London, Surrey, United Kingdom

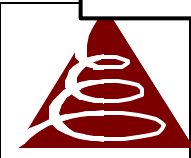
^eInstitute of Molecular Genetics, Academy of Sciences of the Czech Republic, Videňská 1083, 142 20 Prague 4, Czech Republic.

Page 1 Sec 1 1/27 At 5.8" Ln 17 Col 1



Coded File Example

```
<ATL>Nuclear  $\gamma$ -Tubulin during Acentriolar Plant Mitosis</ATL>
<AUG>Pavla Binarova*, Věra Cenklovb, Bettina
  Hausec, Elena Kubatova, Martin
  Lysbkb, Jaroslav Doležbelb, Laszlo
  Bouml;gred, and Pavel Drabere</AUG>
<AFF>aInstitute of Microbiology, Academy of Sciences of the Czech Republic,
  Vedenska; 1083, 142 20 Prague 4, Czech Republic.</AFF>
<AFF>bInstitute of Experimental Botany, Academy of Sciences of the Czech
  Republic, Sokolovska; 6, 772 00, Olomouc, Czech Republic.</AFF>
<AFF>cInstitute of Plant Biochemistry, P.O.Box 110432, D-06018 Halle,
  Germany.</AFF>
<AFF>dSchool of Biological Sciences, Royal Holloway, University of London,
  Surrey, United Kingdom</AFF>
<AFF>eInstitute of Molecular Genetics, Academy of Sciences of the Czech
  Republic, Vedenska; 1083, 142 20 Prague 4, Czech Republic.</AFF>
<COR>*To whom correspondence should be addressed. E-mail
  <UNL>binarova@biomed.cas.cz</UNL>; fax 420-2-4752384.</COR>
<RRH>Running title:  $\gamma$ -Tubulin in Plant Mitosis</RRH>
```



Electronic Workflow, v 1.0

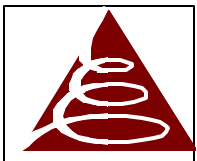
◆ Advantages

- Better than paper
- Avoided SGML tool limitations
- Minimized Training costs



◆ Disadvantages

- Three file conversions
- Error-prone editorial workflow
- Errors discovered in SGML creation



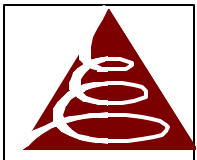
Back to the Drawing Board

◆ XML first

- Convert to XML immediately
- Edit in XML
- Print from XML

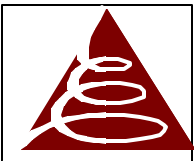
◆ Just in time XML

- Edit in Microsoft Word with styles
- Print from “light weight” XML
- Add granularity when necessary



XML First Workflow

- ◆ Submit electronic manuscript
- ◆ Convert to XML file
- ◆ Edit XML file
- ◆ Typeset from XML
- ◆ Proof and typeset corrections
- ◆ Print and create PDF



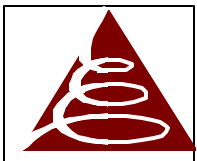
Advantages and Disadvantages

◆ Advantages

- Only one file conversion
- File is continually parsed

◆ Disadvantages

- Tools have just caught up
- Training is expensive
- Editors work amidst XML tags
 - or XML editing customization is expensive



Tiptoeing Through the Tags

◆ Print Version

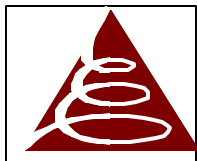
Neutra, R., Shusterman, D. (1991) Hypotheses to explain the higher symptom rates observed around hazardous waste sites. *Environmental Health Perspectives* **94**, 31–38.

◆ XML, DTD 1

```
<bb id="b7"><jnlref>  
  <au><snm>Neutra</snm><x>, </x><fnms>R.</fnms></au><x>, </x>  
  <au><snm>Shusterman</snm><x>, </x><fnms>D.</fnms></au><x> (</x>  
  <cd year="1991">1991</cd><x>) </x><tl>Hypotheses to explain the higher  
  symptom rates observed around hazardous waste sites.</tl><x> </x>  
  <pubtl>Environmental Health Perspectives</pubtl><x> </x><vid>94</vid><x>, </x>  
  <ppf>31</ppf><x>&ndash;</x><ppl>38</ppl><x>.</x></jnlref></bb>
```

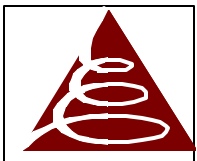
◆ XML, DTD 2

```
<CITATION ID="rf7" URI-STATUS="NORESOLVE" BIB-STATUS="NOLINK">  
  Neutra, R., Shusterman, D. (1991) Hypotheses to explain the higher symptom rates observed  
  around hazardous waste sites. <ITAL>Environmental Health Perspectives</ITAL>  
  <BF>94</BF>, 31 &ndash;38.</CITATION>
```



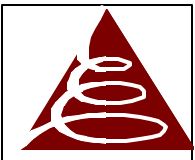
Just-In-Time XML

- ◆ Gradual enrichment
 - Only necessary tags for task at hand
- ◆ Keep editors focused on text
 - Automate tedious tasks
- ◆ Proofing without tagging
 - Use pattern recognition for some tagging
 - Pre-process XML

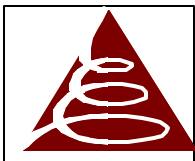
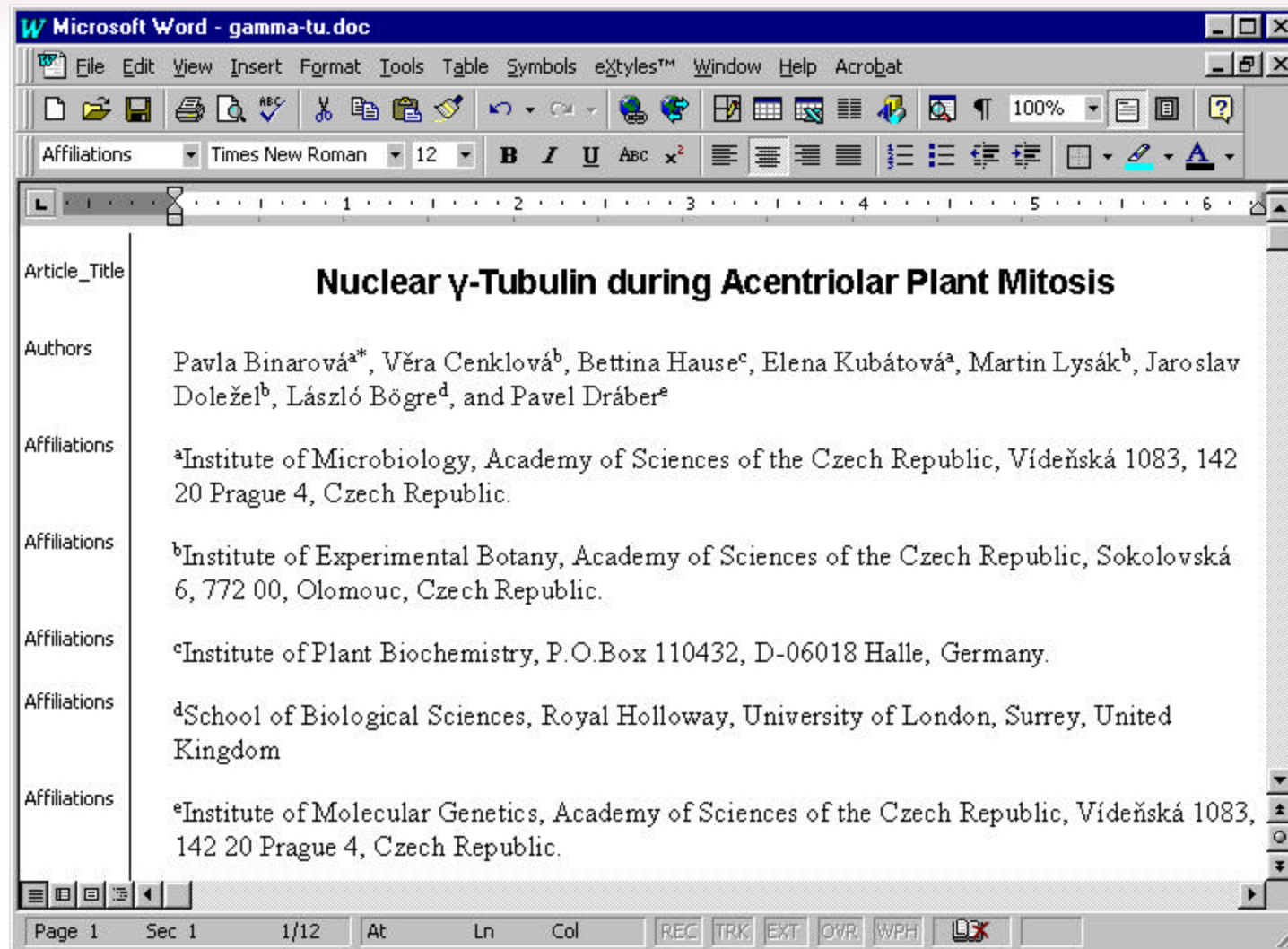


Just-In-Time XML Workflow

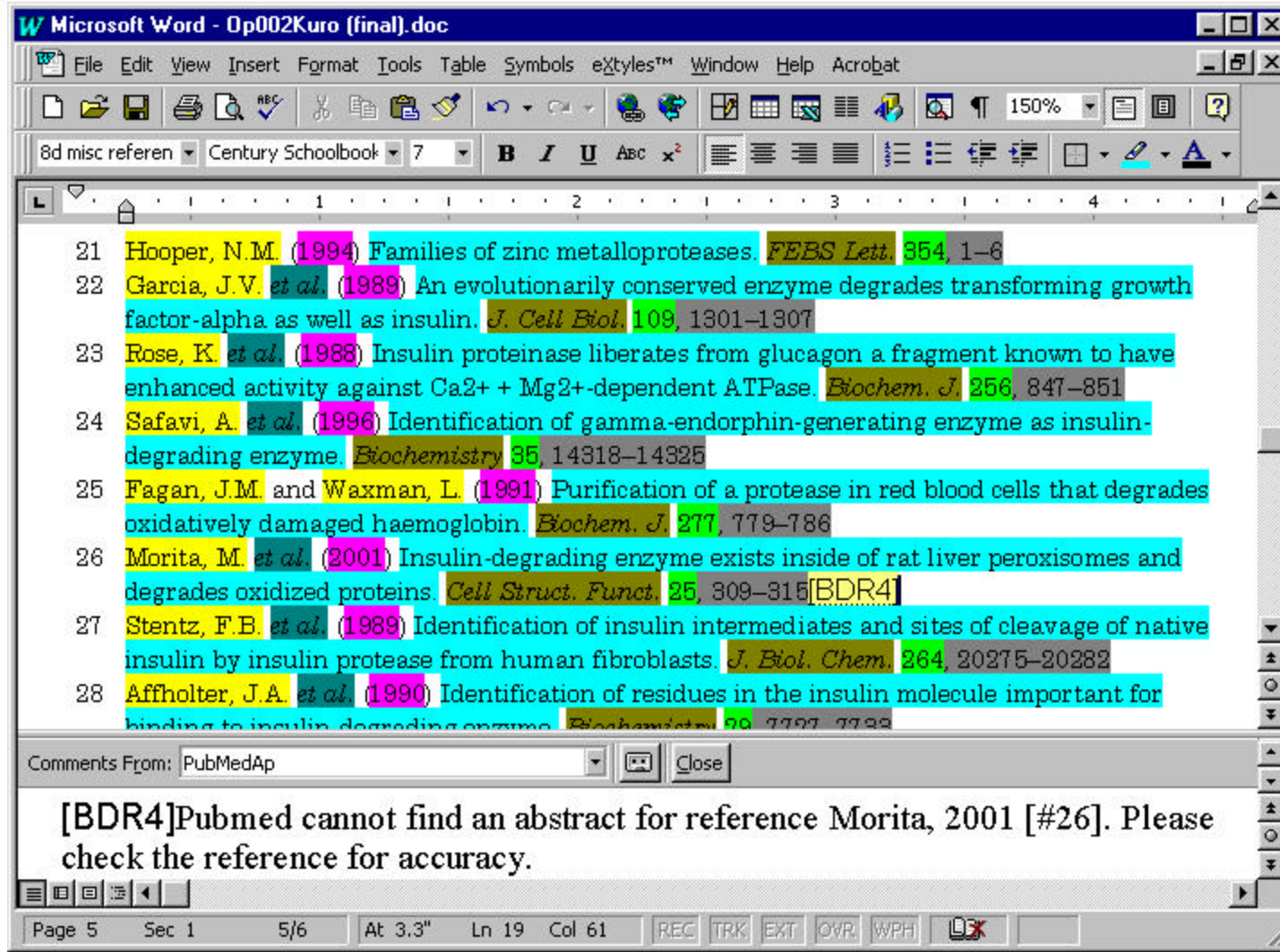
- ◆ Submit electronic manuscript
- ◆ Clean up and style paragraphs
- ◆ Edit in Microsoft Word
- ◆ Typeset from lightweight XML
- ◆ Proof and typeset corrections
- ◆ Enrich XML
- ◆ Print and create PDF



Just-In-Time Editing



Proofing Without Tagging



Microsoft Word - Op002Kuro (final).doc

File Edit View Insert Format Tools Table Symbols eXtyles™ Window Help Acrobat

8d misc referen Century Schoolbook 7 B I U ABC x²

21 Hooper, N.M. (1994) Families of zinc metalloproteases. *FEBS Lett.* 354, 1-6

22 Garcia, J.V. et al. (1989) An evolutionarily conserved enzyme degrades transforming growth factor-alpha as well as insulin. *J. Cell Biol.* 109, 1301-1307

23 Rose, K. et al. (1988) Insulin proteinase liberates from glucagon a fragment known to have enhanced activity against Ca²⁺ + Mg²⁺-dependent ATPase. *Biochem. J.* 256, 847-851

24 Safavi, A. et al. (1996) Identification of gamma-endorphin-generating enzyme as insulin-degrading enzyme. *Biochemistry* 35, 14318-14325

25 Fagan, J.M. and Waxman, L. (1991) Purification of a protease in red blood cells that degrades oxidatively damaged haemoglobin. *Biochem. J.* 277, 779-786

26 Morita, M. et al. (2001) Insulin-degrading enzyme exists inside of rat liver peroxisomes and degrades oxidized proteins. *Cell Struct. Funct.* 25, 309-315[BDR4]

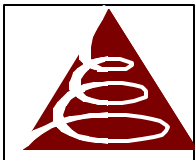
27 Stentz, F.B. et al. (1989) Identification of insulin intermediates and sites of cleavage of native insulin by insulin protease from human fibroblasts. *J. Biol. Chem.* 264, 20275-20282

28 Affholter, J.A. et al. (1990) Identification of residues in the insulin molecule important for binding to insulin-degrading enzyme. *Biochemistry* 29, 7727-7732

Comments From: PubMedAp Close

[BDR4]Pubmed cannot find an abstract for reference Morita, 2001 [#26]. Please check the reference for accuracy.

Page 5 Sec 1 5/6 At 3.3" Ln 19 Col 61 REC TRK EXT OVR WPH



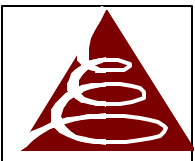
Just-In-Time XML Composition

◆ Lightweight XML citation

```
...public.(<xref>1</xref>-<xref>3</xref>)
```

◆ Final XML citation

```
...public.(<xref ref-type="bibr" rid="R1">1</xref>-<xref  
ref-type="bibr" rid="R3">3</xref>)
```



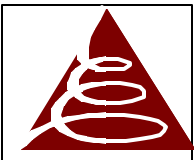
Advantages and Disadvantages

◆ Advantages

- Editors work in Microsoft Word
- Lower training costs
- Freelance editors are practical
- Errors are caught prior to XML creation

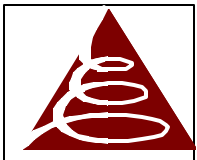
◆ Disadvantages

- Two file conversions
- Structure is enforced later



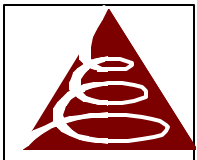
The Benefits

- ◆ **Costs are lower**
 - Copy-editing is faster
 - Typesetting is more accurate
- ◆ **Quality is higher**
 - Print quality is improved
 - XML quality is improved
- ◆ **Production is faster**
 - Content is published sooner

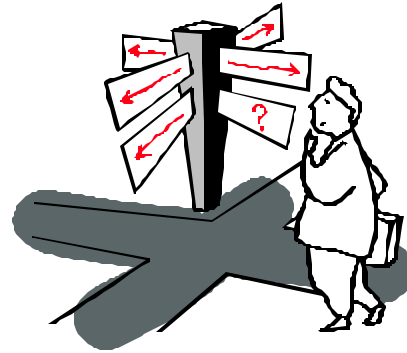


Conclusions

- ◆ XML is not just about tags
- ◆ XML is about new workflows
 - Lower costs
 - Higher quality
 - Faster Production
- ◆ XML is about new opportunities
 - New products
 - New ideas



Questions?



Bruce Rosenblum
Inera Incorporated
+1 (617) 969 - 3053
brosenblum@inera.com
www.inera.com

