



Editorial and XML
Solutions for Publishers

So... What Have We *Really* Been Up To?

Exciting new developments
in Inera technology

**Elizabeth Blake
and Bruce Rosenblum**

Inera Development Projects

- ▶ Juggling priorities
 - ▷ Improving existing features of eXtyles and Edifix
 - ▷ Adding new features to eXtyles and Edifix
 - ▷ Custom development work for new and existing customers (e.g., we're finally tackling DITA!)
 - ▷ Staying ahead of updates from our various technology partners (Microsoft, Crossref, NCBI, MathType, etc.)

- ▶ Looking at the big picture
 - ▷ What are the major challenges in publishing? What will they be in 5 years?
 - ▷ How can we help address them?

Challenges in Scholarly Publishing

- ▶ Open Access
 - ▷ Pressure to make ever more content freely available
 - ▷ Establishing financially sustainable OA models
- ▶ Time to publication
- ▶ Identifying trusted scholarship
- ▶ Quality vs. cost: some publishers can afford to invest more than others in labor-intensive or automated quality measures

When Challenges Converge

- ▶ There's a widespread drive to publish **more content, more quickly**, with **less labor** and **more automation**
- ▶ How to do *all* of this without sacrificing quality?

Inera, Automation, and Quality

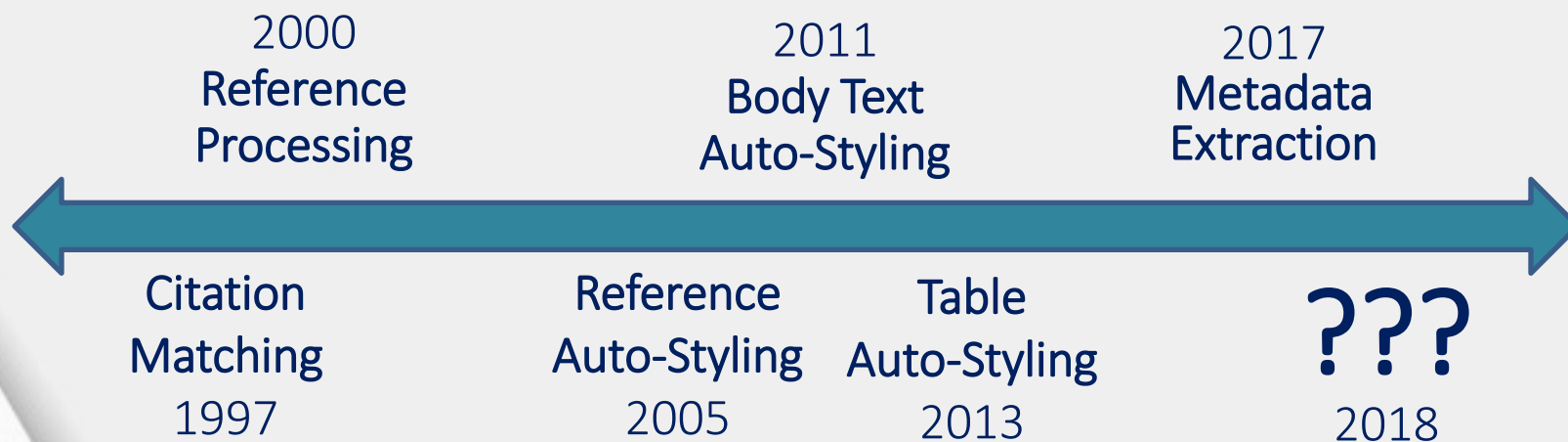
- ▶ Intelligent automation
 - ▷ First do no harm: automate what can be safely automated to save time and enhance consistency and accuracy of content
 - ▷ Leave the stuff that requires human intelligence to the experts
- ▶ Flexibility and scalability
 - ▷ Some of our customers are high touch with their publications; some are very low touch
 - ▷ Some are large, established publishers; some are just getting up and running
 - ▷ Wide variety of content types, workflows, and requirements to which eXtyles adapts

eXtyles Desktop and Server Versions

- ▶ eXtyles (Desktop): 2000
- ▶ eXtyles SI (Server Implementation): 2005
- ▶ SI was originally developed for Aries Systems
 - ▷ Integration of eXtyles reference tools into EM
 - ▷ Required building automatic recognition of the reference list in articles in Word (eXtyles Auto-Style)
- ▶ SI has been adopted by many other organizations for use in combination with eXtyles Desktop to more fully automate certain aspects of the eXtyles workflow

A Brief History of eXtyles Automation

Specifically,
automatic recognition of document elements



What's Next?

The biggest hurdle is successful automatic conversion from author submission to XML

- ▶ XML authoring tools have not been adopted
- ▶ Author compliance with Word templates or other formatting requirements is unreliable
- ▶ JATS is complex
- ▶ Word is complex
- ▶ Authors are complex (and smart, and creative)
- ▶ If your automated tool only works on a perfectly formatted article, then it's not very automated

What's Next?

- ▶ Inera has continually developed new features for automatic recognition of document elements within eXtyles
- ▶ One eXtyles process has always required manual labor: Paragraph Styling
- ▶ The next frontier: fully automated document structuring
- ▶ Introducing...

eXtyles Arc

Automated markup of unstructured content

eXtyles Arc

Automated markup of unstructured content

- ▶ Extracts metadata
- ▶ **Automatically applies paragraph styles**
- ▶ Runs all reference and citation processes
- ▶ Logs inconsistencies and errors
- ▶ Exports to JATS
- ▶ Can be a fully automated publishing solution or used in combination with high-touch editorial workflows

Current Status

- ▶ eXtyles Arc release 1.0:
 - ▷ Journal article content
 - ▷ Available with eXtyles SI
 - ▷ Focused on STM content elements

- ▶ Initial Goal:
 - ▷ 100% accurate JATS from 80% of submissions
 - ▷ The other 20% will require touch-up of the XML, but there will be no loss of content

eXtyles Improvements

- ▶ The eXtyles Arc project required improvements that have already cascaded down to existing releases of eXtyles Desktop and SI
- ▶ Without human intervention, the things that have already been automated for a long time had to get even better
- ▶ Other improvements are likely to follow

eXtyles Desktop/Sl Improvements

- ▶ Major improvements to Author Processing (required for Metadata Extraction)
- ▶ XML export of table titles/captions placed below, rather than above, the table
- ▶ Minimization of certain classes of XML parsing errors
 - ▷ Author footnote symbols with no corresponding footnote
 - ▷ Uncited tables
 - ▷ In a nutshell: **more user-friendly dialog warnings instead of cryptic XML errors**

Possible Future eXtyle Improvements

eXtyle Arc innovations that have not yet been added to current releases but may be available in the future include:

- ▶ Detection of duplicate content (e.g., an article title appears twice in the Word file)
- ▶ Automatic recognition of the in-text citation style used by the author
- ▶ And, obviously, full Auto-Styling of content (no immediate plans to implement this in the Desktop environment, but we want to hear your feedback)

eXtyles Arc in Action! (Demo)

**Questions?
Comments?
Grand visions?**

Speak now or send them to Liz at
lblake@inera.com