A complete end-to-end publishing system based on JATS

Kaveh Bazargan
Brief history

- Physicist (3D imaging and Holography)
  - First TeX user in London University (1983)
- Started company in London in 1988
- Offices in UK, India, Europe
<table>
<thead>
<tr>
<th>Our Clients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IOP</strong> Institute of Physics</td>
</tr>
<tr>
<td><strong>npg</strong> nature publishing group</td>
</tr>
<tr>
<td><strong>OXFORD UNIVERSITY PRESS</strong></td>
</tr>
<tr>
<td><strong>CASPA</strong></td>
</tr>
<tr>
<td><strong>QSCIENCE.com</strong></td>
</tr>
<tr>
<td><strong>WILEY-BLACKWELL</strong></td>
</tr>
</tbody>
</table>
Our Business

• Typesetting of academic books and journals
  • For 15 years, using true XML-first PDF output
• Last 8 years, platforms for publishers
Our philosophy

- Automate where possible
- Use free open software
- Nothing should be typed in twice!
- We should not be in business!
- Keep workflows simple
Why free software

- No software costs!
- Lower prices for clients
- No dependency on other companies – inc Adobe
- Can extend software
- Can do what commercial software can’t do
Example: TeX/LaTeX

• True automated JATS to PDF
• Highest typographic control
• For anything except page by page graphic design
• Multilingual footnotes, marginal notes, indexes
• Allows automated enhanced PDFs from XML
• Demo…
The publication process

- Authoring
- Editorial
  - Submission and peer review
- Production
  - XML, PDF, etc – *must* be equivalent
The authoring problem
MacWrite

MacWrite is a leading word processing application. It can be used to write memos, reports, etc. Graphics and pictures from other applications can be pasted into MacWrite. In addition, words can be emphasized by changing styles or changing sizes.

The pull down menus across the top contain all the commands you will ever need. If you can cut, copy, and paste, you can use MacWrite. For instance, to replace one word with another word simply select the item named “Change” from the “Search” Menu.
The word processor

• Good if document is final product
• But it is page-based…
• Writer is encouraged to create a nice page
• Too much freedom to change document
Ingredients of a typical manuscript

- Title
- Authors and affiliations
- Abstract
- Keywords
- Sections/subsections
- Citations
- References
- Acknowledgments
- Funding statement …
Made to look good visually
An analogy...
Now get the fruit back!

• This is the present workflow in publishing
• Authors create manuscript, visually
• But publishers want XML – i.e. ingredients
Example: references
Authors use reference managers

- EndNote
- Mendeley
- Zotero
A Novel Finding of Sentinel Lymphatic Channels in Early Stage Breast Cancer Patients: Which May Influence Detection Rate and False-Negative Rate of Sentinel Lymph Node Biopsy

Minghai Wang1,2#, Wenbin Zhou1#, Yingchun Zhao1#, Tiansong Xia1, Xiaoming Zha1, Qiang Ding1,
Xiaoan Liu1, Yi Zhao1, Lijun Ling1, Lin Chen1, Shui Wang1*

1 Department of Breast Surgery, The First Affiliated Hospital with Nanjing Medical University, Nanjing, Jiangsu, China, 2 Department of General Surgery, The First Affiliated Yijishan Hospital with Wannan Medical College, Wuhu, Anhui, China.

To add a note, highlight some text. Hide notes

Jump to
Abstract
Introduction
Patients and Methods
Results
Title: A Novel Finding of Sentinel Lymphatic Channels in Early Stage Breast Cancer Patients: Which May Influence Detection Rate and False-Negative Rate of Sentinel Lymph Node Biopsy

Authors: Wang, Minghai, Zhou, Wenbin, Zhao, Yingchun, Xia, Tiansong, Zha, Xiaoming, Ding, Qiang, Liu, Xiaohan, Zhao, Yi, Ling, Lijun, Chen, Lin, Wang, Shui

Abstract: Background The exact lymphatic draining route remains unclear in some cases of early-stage breast cancer. This study aimed to identify sentinel lymphatic channels (SLCs) at the level of axillary lymph nodes (ALN) or in the perivascular region around the primary tumor. Methods A total of 234 breast cancer patients underwent sentinel lymph node biopsy for ALN before surgery. SLCs were identified using a fluorescent lymphatic mapping agent (LMA) and confirmed by histology. Results We identified 234 SLCs in 234 patients (99.6%). The SLCs were classified into three types: type A (20.7%), type B (20.7%), and type C (58.6%). Type A SLCs were smaller and more difficult to identify than type B and C SLCs. Type A SLCs were more commonly detected in patients with larger tumors than type B and C SLCs. Type A SLCs were also more commonly detected in patients with high histological grade tumors than type B and C SLCs. Conclusions Our study provides valuable insights into the anatomy and function of SLCs in early-stage breast cancer patients, which may influence the detection rate and false-negative rate of sentinel lymph node biopsy.

Publication: PLoS ONE

Volume: 7
Issue: 12
Pages: e51226
Date: December 4, 2012

DOI: 10.1371/journal.pone.0051226
ISSN:
Author instructions (1000s)

• Authors asked for style of journal
  • Order of ingredients (Year, issue, page)
  • Punctuation
• Typeface
• Journal abbreviations…
Ref manager exports correct style

- Correct punctuation…
- Typeface
- Order of “ingredients”
- Author thinks they are helping the process…
- but they are not!
File goes to typesetter

- Publisher requires full XML
- And (rightly) XML first workflow
- Typesetting reverse engineers references…
- to get full reference data, e.g. CrossRef
But we need PDF to style

• Typesetter writes a filter to convert XML of reference to journal style once more
Current industry practice

• Follow majority of authors, so build clever Word-based tools, to extract ingredients

• Still need a quick check, cannot automate

• Manual work increases with complexity of doc
A better way?
How a paper is published now…
Advantages of XML-centric workflow

• Fast and low cost
• Perfect, structured XML throughout
• Truly future-proof
• Accessible files for blind, dyslexic, etc
Requirements

• Blindingly fast

• Only XML is saved in system

• PDF pagination fixed, even 20 years hence

• PDFs etc created on the fly, as required by reader
  • Customize, e.g. color/black and white
  • Embed info, e.g. comments, peer review data
Transcriptome analyses of *Anguillicola crassus* from native and novel hosts

Parasite is native to the

European eel

species and the European eel

A. crassus has been proposed for this introduction. In the new host in the recipient area, the parasite appears to be more pathogenic. As a reason for these differences, genetically fixed differences in infectivity and development between Taiwanese and European A. crassus have been described and disentangled from plasticity induced by different host environments. To explore whether transcriptional regulation is involved in these lifecycle differences, we have analysed a “common garden”, cross infection experiment, using deep-sequencing
RVPublisher modules

- **RVRite** – authoring
- **ReView** – user-friendly peer review
- **RVEdit** – copy editing
- **ProofCheck** – author proof correction
- **RVFormatter** – Auto-conversion to PDF, Epub, …
RVRite version 1
Schlager Group

Schlager Group Production

Blues
Author: D. Kelly
0 Comments

Hewis
Author: D. Kelly
0 Comments

First
Author: M. Holmes
0 Comments

Lovea
Author: A. Paris
0 Comments

Whata
Author: C. Dominic
0 Comments

Chine
Author: A. Miller
0 Comments

Cordx
Author: K. Greenberg
0 Comments

Night
Author: B. Skeen
0 Comments

Whenl
Author: B. Aubrey
0 Comments
A complex document

Zoë Akins’s *The Old Maid* is an adaptation of Edith Wharton’s novel by the same name.
Output is clean and readable

1935

Zoë Akins's *The Old Maid* is an adaptation of Edith Wharton’s novel by the same name. The five-episode melodrama takes place in New York City in the nineteenth century. Set in a time when Victorian sensibilities determined a woman’s role and choices in life, the play explores the themes of jealousy, love, class, and sacrifice. Produced in 1935, *The Old Maid* won the Pulitzer Prize for drama and ran for two years. The play was so successful that it was adapted for a film that was released in 1939.

First published in 1935, *The Old Maid* is no longer in print, but an eBook version is available on Questia. This book contains racially charged language in one episode that some readers may find offensive.

**Author Biography**

Zoë Akins was born in Humansville, Missouri, on October 30, 1886. Her family moved to St. Louis in 1898 when her father, Thomas Jasper Akins, was made the State Chairman for the Republican Party. She graduated from high school in 1903 and was engaged to Marion Reed when she was seventeen. Reed was the editor
Author queries

1800s, when independence movements in South America broke Spain's hold on the New World. In 1898 the United States defeated Spain in the Spanish–American War, driving it out of its last colonies in Cuba and Puerto Rico and thus ending Spanish rule in the Americas.

11 Responses to "Battle of Tenochtitlan (1521)"

marcia
December 22, 2010 at 8:02 pm  Quote
[A] A bit of a stretch perhaps. His papal bull set a demarcating line for lands claimed by Spain and those by Portugal. The line was later shifted with the Treaty of Tordesillas in 1494. Portugal, after all, had Brazil.
Edit  Reply

M_oneal
December 23, 2010 at 4:10 pm  Quote
Let’s just delete the sentence. It’s not critical, and it leads to the need for explanation.
Edit  Reply

marcia
December 22, 2010 at 8:02 pm  Quote
[A] Note addition of "June" to clarify later chronology.
Edit  Reply
XML
John Gardner – ViewPlus
Accessibility is the litmus test for good XML

- You need good, structured XML
- 100% confidence in content…
- …more confidence than in the PDF
- PDF is just one rendering
- So use XML-first workflow
Parsing is not enough
I am the only one looking at XML!
A simple, common error

\begin{verbatim}
(year>2005</year>) ... <volume>5</volume>:
<fpage>341</fpage><lpage>354</lpage>
\end{verbatim}

\begin{verbatim}
<fpage>341</fpage>
<lpage>54</lpage>
\end{verbatim}
species in having pales, 0.06±0.01 in females, 0.06±0.01 in females, 0.06±0.01 in females, 0.06±0.01 in females, 0.06±0.01 in females.

ionally smaller tympana (TY/SVL = 0.06<underline>+</underline>0.01 in both
we redefined precision as follows:

\[
\frac{TPP}{TPP + FPP} = \frac{TPP \text{ for TFs with ChIP data}}{TPP \text{ for TFs with ChIP data} + FPP \text{ for TFs with ChIP data}}
\]

\[
= \frac{TPP \text{ for TFs with ChIP data}}{All \ predictions \ made \ for \ TFs \ with \ ChIP \ data}
\] (10)

Recall was calculated as previously described [43]:

\[
\frac{TPP}{TP} = \frac{True \ positive \ predictions}{All \ known \ true \ positives}
\]

(11)
The answer...

- True automated XML-first
- PDF can be deleted and recreated at any point
- Demo...