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

Kaveh Bazargan

RIVER VALLEY

www.RiverValleyTechnologies.com @kaveh1000

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





Our Clients

IOP Institute of Physics	ELSEVIER
npg nature publishing group	CAMBRIDGE UNIVERSITY PRESS
OXFORD UNIVERSITY PRESS	SPIE.
CASPA	stm
QSCIENCE.com	PeerJ
WILEY-BLACKWELL	

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

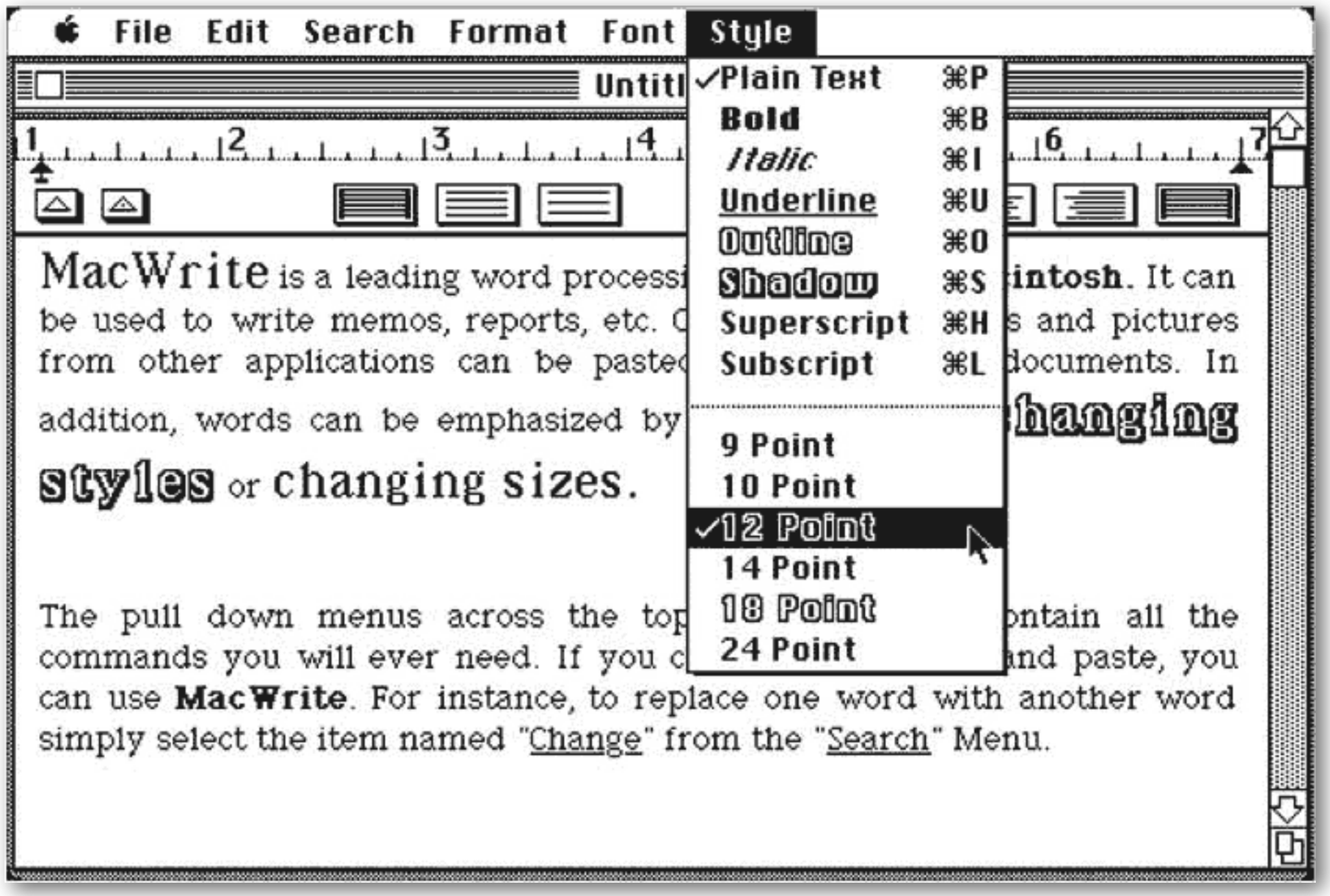
- True automated JATS to PDF
- Highest typographic control
- For anything except page by page graphic design
- Multilingual footnotes, marginal notes, indexes
- Allows automated *enhanced* PDFs from XML
- Demo…

The publication process

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

The authoring problem

MacWrite


File Edit Search Format Font Style
Untitl
✓Plain Text ⌘P
Bold ⌘B
Italic ⌘I
Underline ⌘U
Outline ⌘O
Shadow ⌘S
Superscript ⌘H
Subscript ⌘L
9 Point
10 Point
✓12 Point
14 Point
18 Point
24 Point
MacWrite is a leading word processi
be used to write memos, reports, etc.
from other applications can be paste
addition, words can be emphasized by
styles or changing sizes.
intosh. It can
s and pictures
documents. In
hanging
The pull down menus across the top
ontain all the
commands you will ever need. If you c
nd paste, you
can use MacWrite. For instance, to replace one word with another word
simply select the item named "Change" from the "Search" Menu.

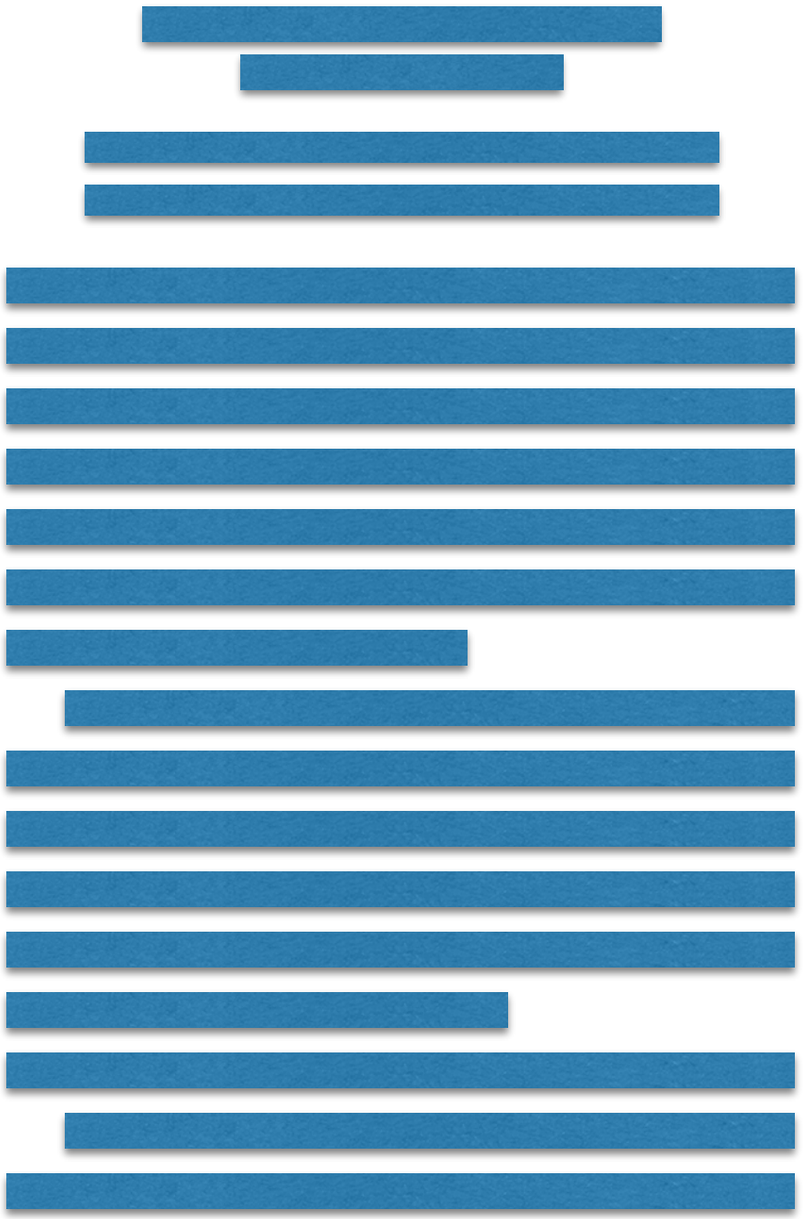
The word processor

- 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

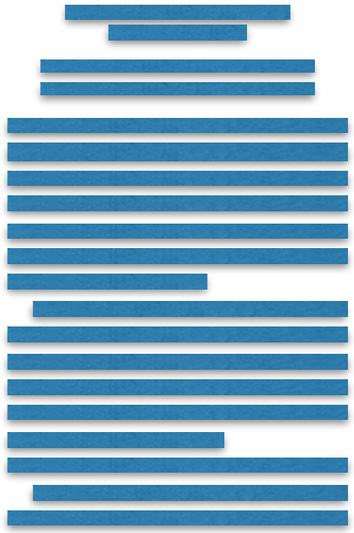
- Title
- Authors and affiliations
- Abstract
- Keywords
- Sections/subsections
- Citations
- References
- Acknowledgments
- Funding statement …

Made to look good visually



An analogy…

Title
Authors
Abstract
Keywords



Pages

Read





Smoothie

Drink

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

www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0051226

PLOS ONE | A peer-reviewed, open access journal

Search articles...

Home Browse Articles About For Readers For Authors and Reviewers

RESEARCH ARTICLE OPEN ACCESS

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

Article Metrics Related Content Comments: 0

Minghai Wang[1,2#], Wenbin Zhou[1#], Yingchun Zhao[1#], Tiansong Xia[1], Xiaoming Zha[1], Qiang Ding[1], Xiaoan Liu[1], Yi Zhao[1], Lijun Ling[1], Lin Chen[1], Shui Wang[1*]

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

www.plosone.org/home.action

To **add a note**, highlight some text. Hide notes

Make a general comment

Jump to

Abstract

Introduction

Patients and Methods

Results

Download:

Print article

EzReprint Ne

Metrics

Recently publi

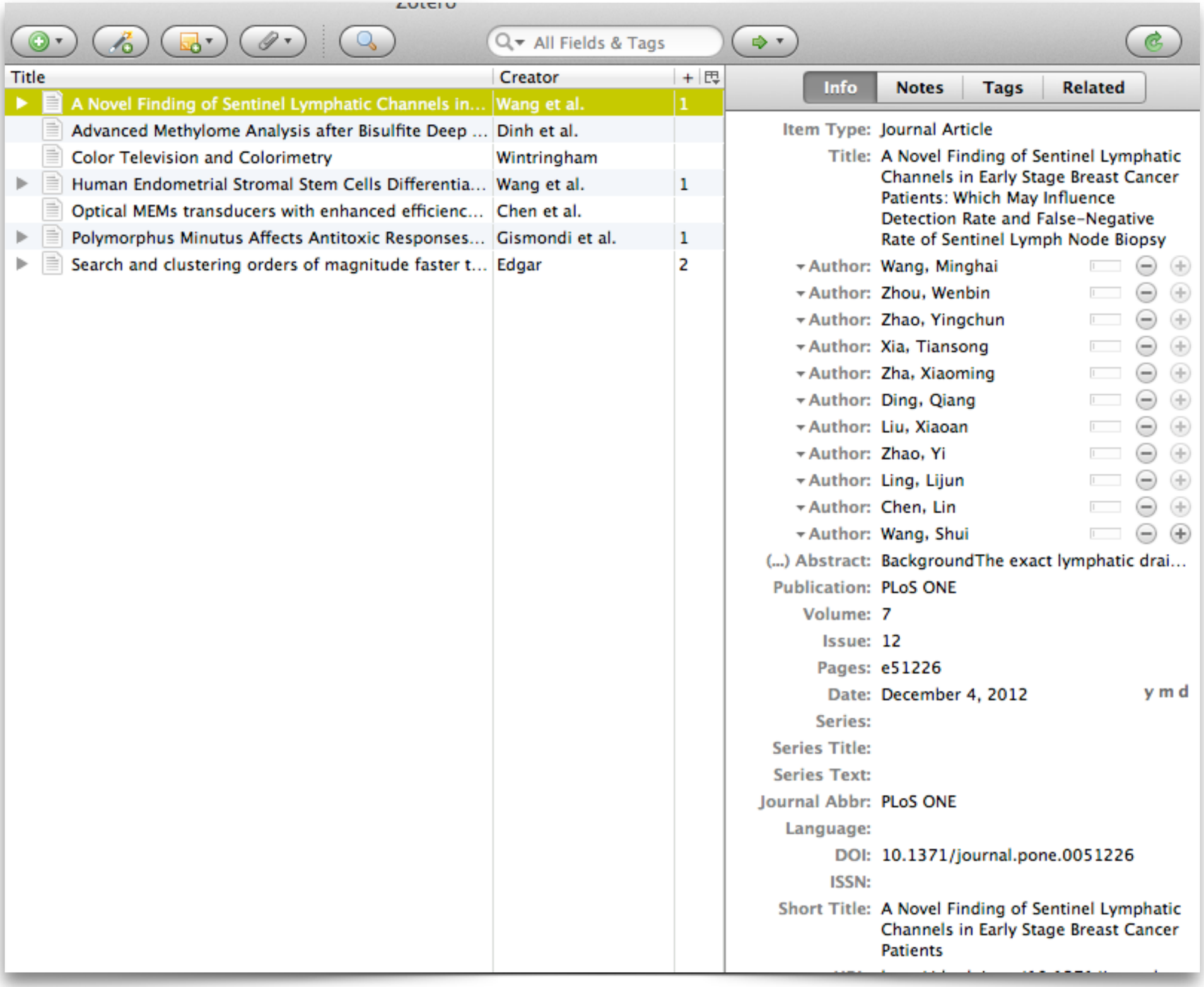
Related Conten

Related Subject

Biology and life sciences, Physica analysis method

Related Articles

Google Scholar


All Fields & Tags
Title
Creator
A Novel Finding of Sentinel Lymphatic Channels in... Wang et al. 1
Advanced Methylome Analysis after Bisulfite Deep ... Dinh et al.
Color Television and Colorimetry Wintringham
Human Endometrial Stromal Stem Cells Differentia... Wang et al. 1
Optical MEMs transducers with enhanced efficienc... Chen et al.
Polymorphus Minutus Affects Antitoxic Responses... Gismondi et al. 1
Search and clustering orders of magnitude faster t... Edgar 2
Info
Notes
Tags
Related
Item Type: Journal Article
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
Author: Wang, Minghai
Author: Zhou, Wenbin
Author: Zhao, Yingchun
Author: Xia, Tiansong
Author: Zha, Xiaoming
Author: Ding, Qiang
Author: Liu, Xiaoan
Author: Zhao, Yi
Author: Ling, Lijun
Author: Chen, Lin
Author: Wang, Shui
(...) Abstract: BackgroundThe exact lymphatic drai...
Publication: PLoS ONE
Volume: 7
Issue: 12
Pages: e51226
Date: December 4, 2012
y m d
Series:
Series Title:
Series Text:
Journal Abbr: PLoS ONE
Language:
DOI: 10.1371/journal.pone.0051226
ISSN:
Short Title: A Novel Finding of Sentinel Lymphatic Channels in Early Stage Breast Cancer Patients

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

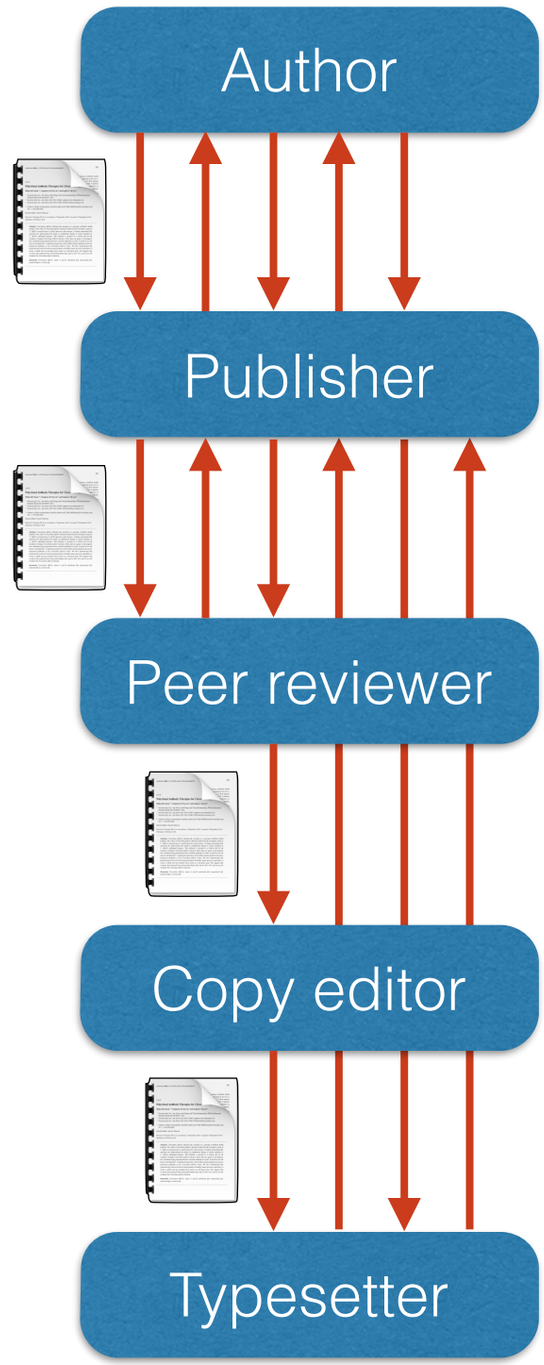
- Typesetter writes a filter to convert XML of reference to journal style once more

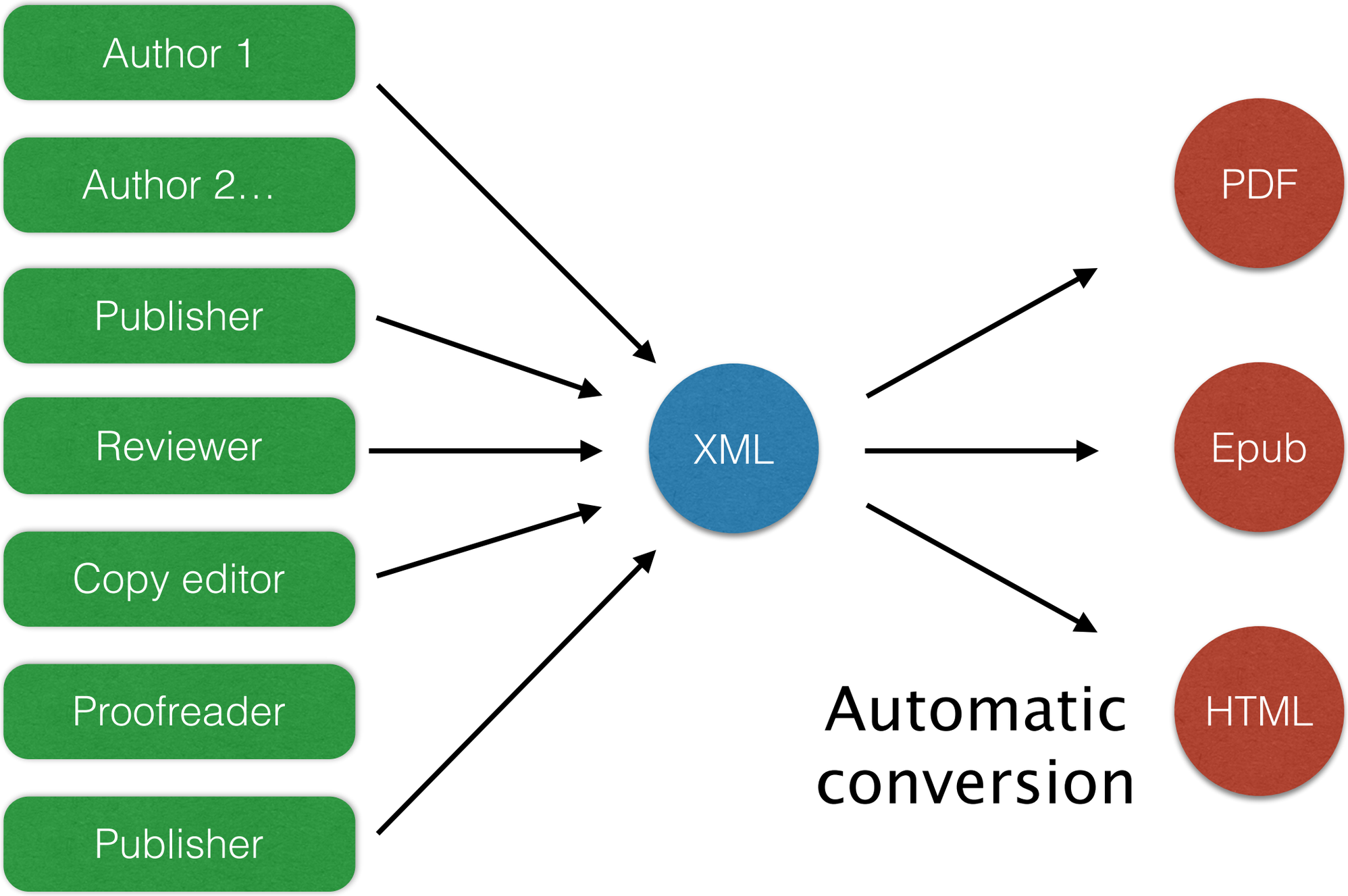
Current industry practice

- 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…


Author
Publisher
Peer reviewer
Copy editor
Typesetter


Author 1
Author 2…
Publisher
Reviewer
Copy editor
Proofreader
Publisher
XML
PDF
Epub
HTML
Automatic conversion

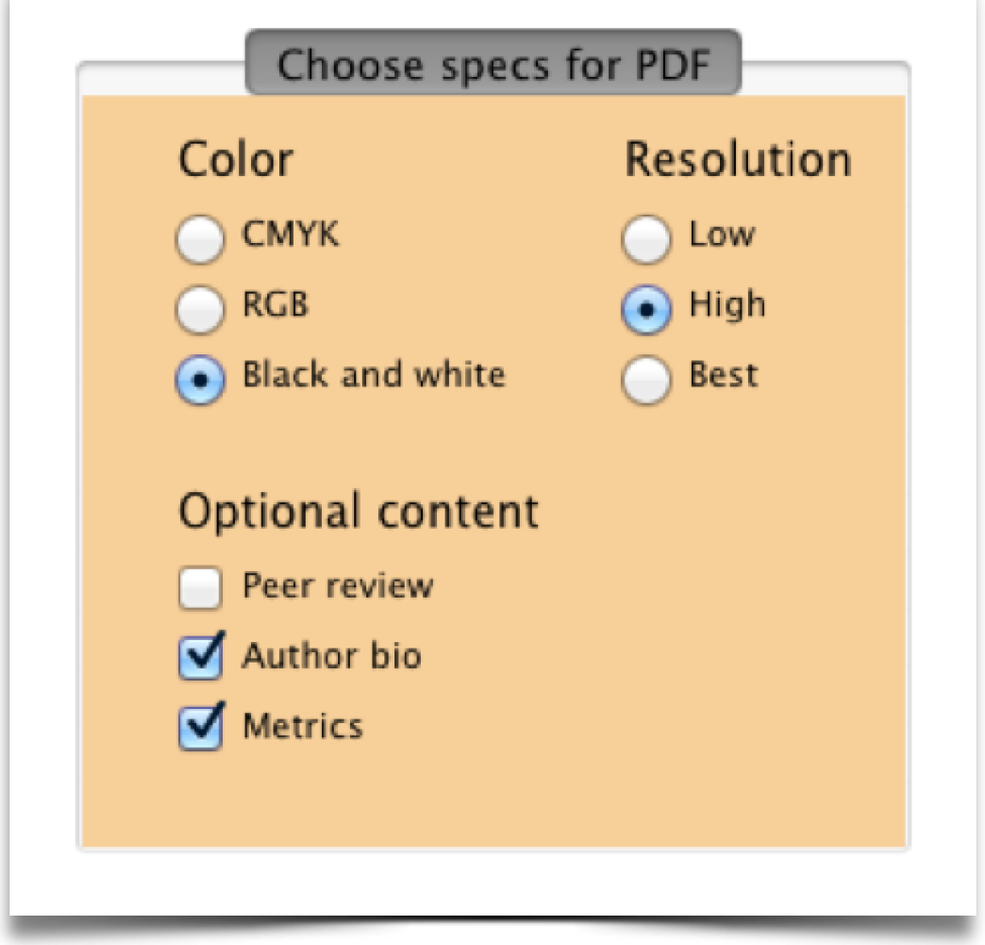
Advantages of XML-centric workflow

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

OpenDyslexic

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


Choose specs for PDF
Color
CMYK
RGB
Black and white
Resolution
Low
High
Best
Optional content
Peer review
Author bio
Metrics

PeerJ

Show data

Metrics
About the author
Reviews
Comments

Transcriptome analyses of *Anguillicola crassus* from native and novel hosts

Usage since published - updated daily

Social referrals unique visitors

Twitter	26
Google+	0
Facebook	86
LinkedIn	0
Reddit	0
Slashdot	0

Social networks

Tweet 14
Recommend 42
Share 0

Top referrals unique visitors

dml.cmnh.org	25
From bookmark or typed URL	21

Close

ɣski[2],

ersity Berlin,

ute for

gh,

of Edinburgh,

arasite is native to the

d the European eel

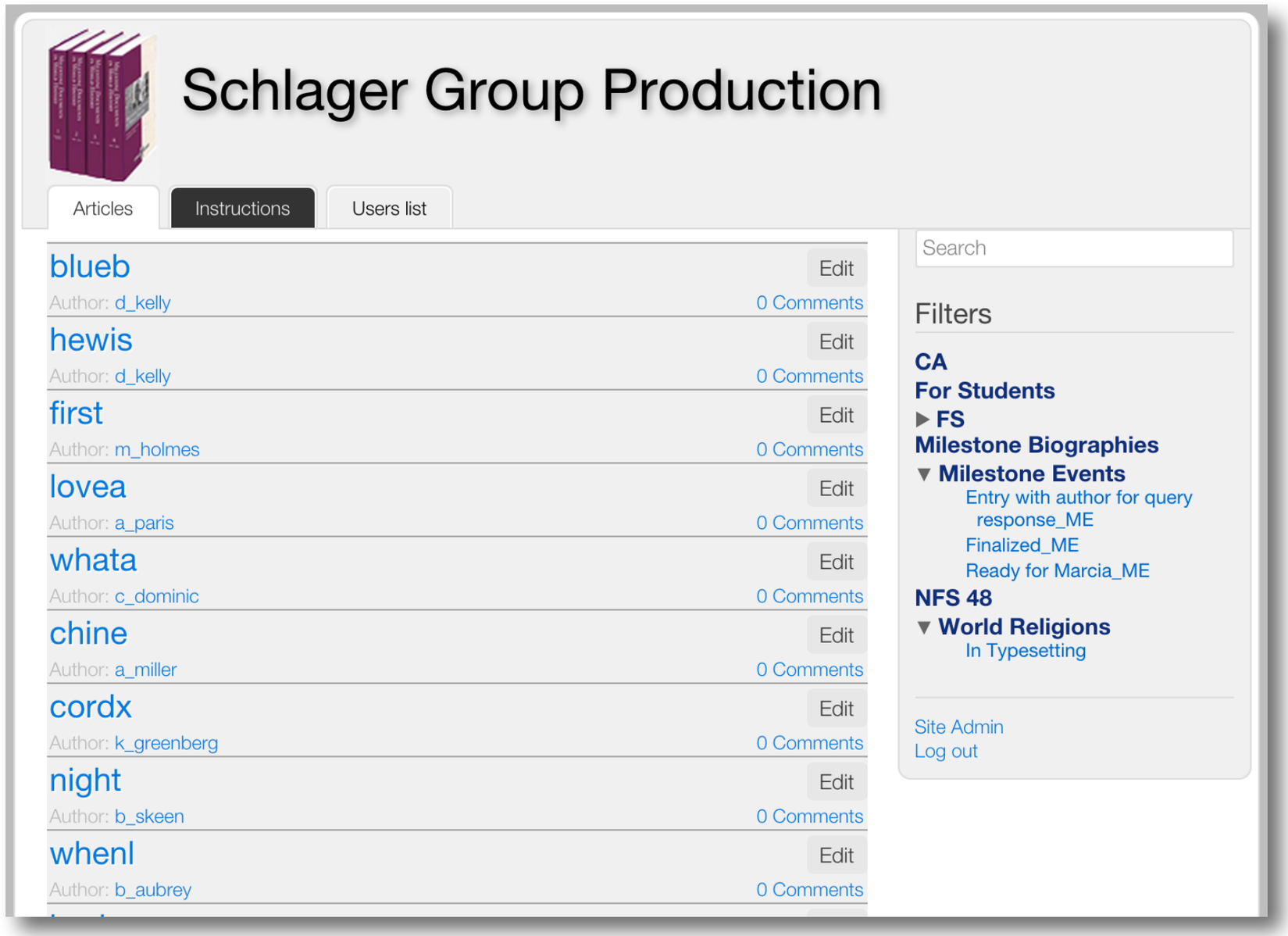
been proposed for this introduction. In the new host in the recipient area, the parasite appears to be more pathogenic. As a reason for these differences, genetically fixed differences in infectivity and development between Taiwanese and European *A.crassus* have been described and disentangled from 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

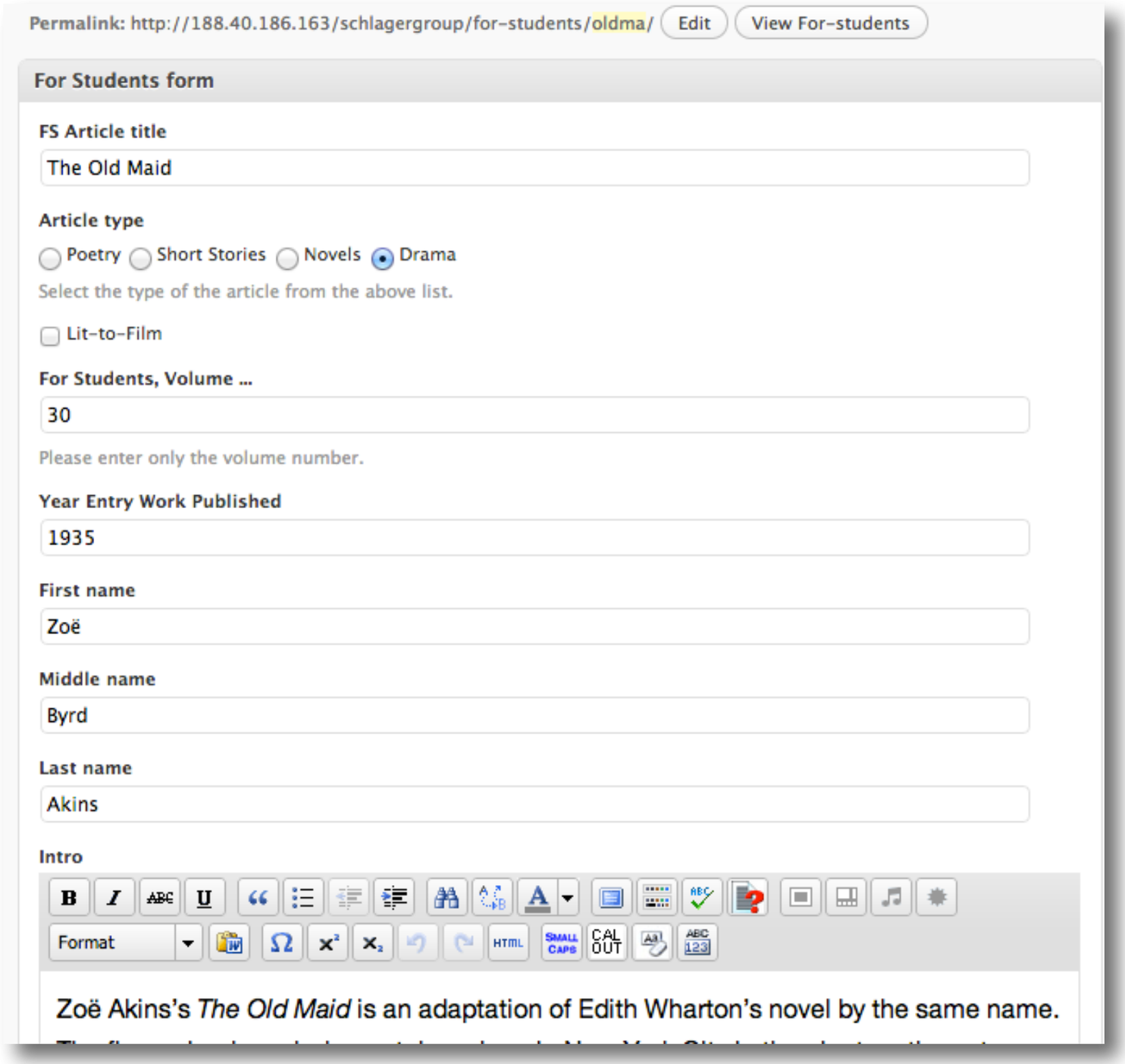
- **RVRite** – authoring
- **ReView** – user-friendly peer review
- **RVEdit** – copy editing
- **ProofCheck** – author proof correction
- **RVFormatter** – Auto-conversion to PDF, Epub, …

RVRite version 1

Schlager Group



A complex document



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 Reedy when she was seventeen. Reedy 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

[auq_1]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

[auq_1]Note addition of "June" to clarify later chronology.

Edit Reply

XML

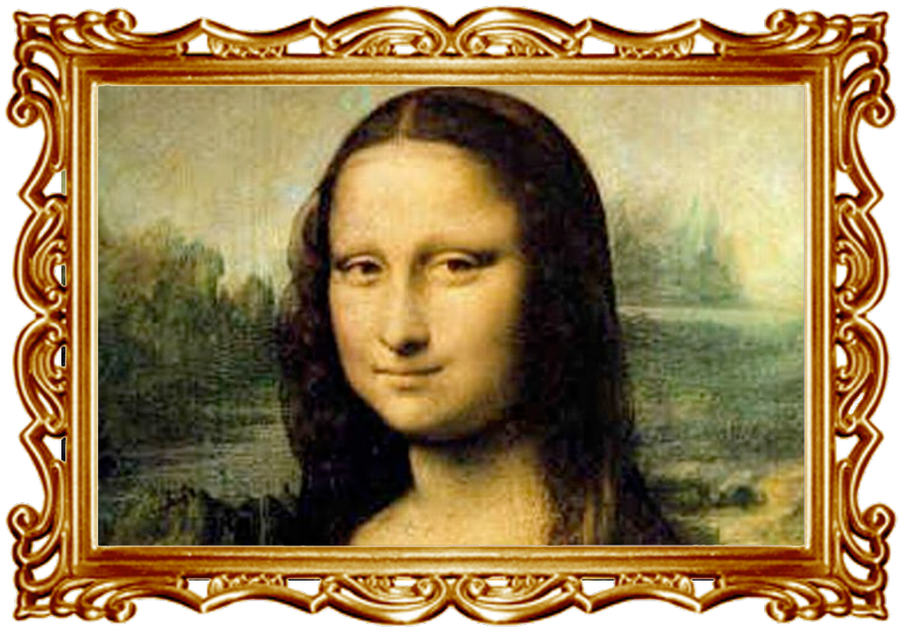
John Gardner – *ViewPlus*



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



√



√

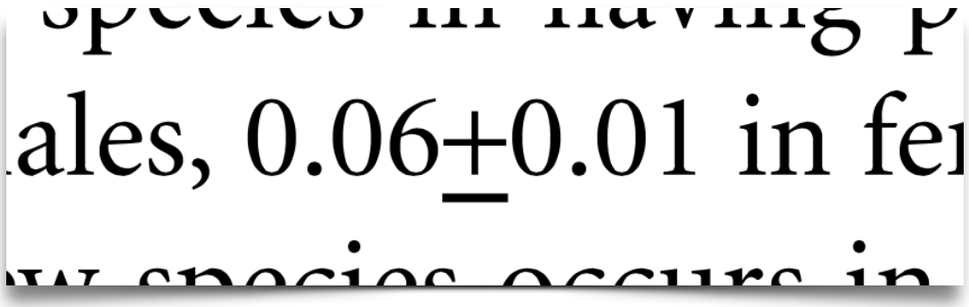
I am the only one looking at XML!

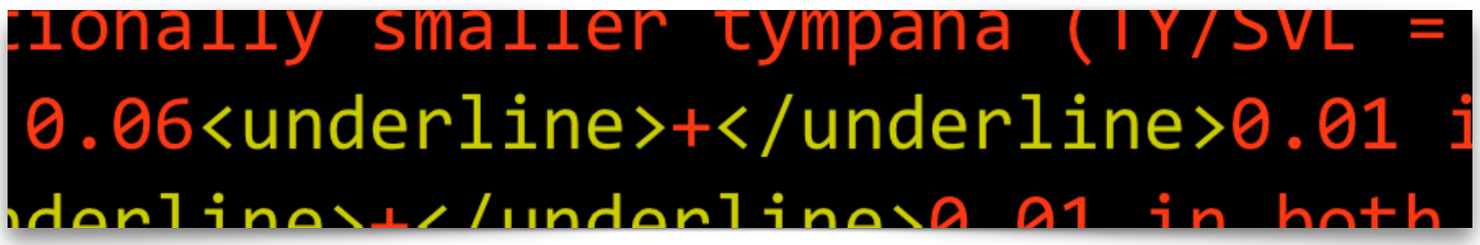
A simple, common error

```
(<year>2005</year>) ... <volume>5</volume>:

<fpage>341</fpage>-<lpage>354</lpage>
```

```
<fpage>341</fpage>
<lpage>54</lpage>
```


ales, 0.06±0.01 in fe


0.06<underline>+</underline>0.01 i

we redefined precision as follows:

$$\frac{TPP}{TPP + FPP} = \frac{TPP\ for\ TFs\ with\ ChIP\ daTa}{TPP\ for\ TFs\ with\ ChIP\ daTa + FPP\ for\ TFs\ with\ ChIP\ daTa}$$
$$= \frac{TPP\ for\ TFs\ with\ ChIP\ daTa}{All\ predictions\ made\ for\ TFs\ with\ ChIP\ daTa} \tag{10}$$

Recall was calculated as previously described [43]:

$$\frac{TPP}{TP} = \frac{True\ positive\ predictions}{All\ known\ true\ positives} \tag{11}$$

www.RiverValleyTechnologies.com @kaveh1000

```
<mml:math display="block" id="M11"
overflow="scroll"><mml:mrow><mml:mfrac><mml:mrow><mml:mi>T</mml:mi><mml:mi>P</
mml:mi><mml:mi>P</mml:mi></mml:mrow><mml:mrow><mml:mi>T</mml:mi><mml:mi>P</mml:
mi></mml:mrow></mml:mfrac><mml:mo>=</mml:mo><mml:mfrac><mml:mrow><mml:mi>T</mml:
mi><mml:mi>r</mml:mi><mml:mi>u</mml:mi><mml:mi>e</mml:mi><mml:mtext> </mml:mtext
><mml:mi>p</mml:mi><mml:mi>o</mml:mi><mml:mi>s</mml:mi><mml:mi>i</mml:mi><mml:mi
>t</mml:mi><mml:mi>i</mml:mi><mml:mi>v</mml:mi><mml:mi>e</mml:mi><mml:mtext> </
mml:mtext><mml:mi>p</mml:mi><mml:mi>r</mml:mi><mml:mi>e</mml:mi><mml:mi>d</mml:
mi><mml:mi>i</mml:mi><mml:mi>c</mml:mi><mml:mi>t</mml:mi><mml:mi>i</mml:mi><mml:
mi>o</mml:mi><mml:mi>n</mml:mi><mml:mi>s</mml:mi></mml:mrow><mml:mrow><mml:mi>A<
/mml:mi><mml:mi>l</mml:mi><mml:mi>l</mml:mi><mml:mtext> </mml:mtext><mml:mi>k</
mml:mi><mml:mi>n</mml:mi><mml:mi>o</mml:mi><mml:mi>w</mml:mi><mml:mi>n</mml:mi><
mml:mtext> </mml:mtext><mml:mi>t</mml:mi><mml:mi>r</mml:mi><mml:mi>u</mml:mi><
mml:mi>e</mml:mi><mml:mtext> </mml:mtext><mml:mi>p</mml:mi><mml:mi>o</mml:mi><
mml:mi>s</mml:mi><mml:mi>i</mml:mi><mml:mi>t</mml:mi><mml:mi>i</mml:mi><mml:mi>v
</mml:mi><mml:mi>e</mml:mi><mml:mi>s</mml:mi><mml:mtext> </mml:mtext></"#6D7376"
mml:mrow></mml:mfrac></mml:mrow>
</mml:math>
```

```
<mml:math display="block" id="M11"
overflow="scroll"><mml:mrow><mml:mfrac><mml:mrow><mml:mi>T</mml:mi><mml:mi>P</
mml:mi></mml:mrow></mml:mfrac><mml:mo>=</mml:mo><mml:mfrac><mml:mrow><mml:mi>T</mml:
mi><mml:mi>r</mml:mi><mml:mi>u</mml:mi><mml:mi>e</mml:mi><mml:mtext> </mml:mtext
><mml:mi>p</mml:mi><mml:mi>o</mml:mi><mml:mi>s</mml:mi><mml:mi>i</mml:mi><mml:mi
>t</mml:mi><mml:mi>i</mml:mi><mml:mi>v</mml:mi><mml:mi>e</mml:mi><mml:mtext> </
mml:mtext><mml:mi>p</mml:mi><mml:mi>r</mml:mi><mml:mi>e</mml:mi><mml:mi>d</mml:
mi><mml:mi>i</mml:mi><mml:mi>c</mml:mi><mml:mi>t</mml:mi><mml:mi>i</mml:mi><mml:
mi>o</mml:mi><mml:mi>n</mml:mi><mml:mi>s</mml:mi></mml:mrow><mml:mrow><mml:mi>A<
/mml:mi><mml:mi>l</mml:mi><mml:mi>l</mml:mi><mml:mtext> </mml:mtext><mml:mi>k</
mml:mi><mml:mi>n</mml:mi><mml:mi>o</mml:mi><mml:mi>w</mml:mi><mml:mi>n</mml:mi><
mml:mtext> </mml:mtext><mml:mi>t</mml:mi><mml:mi>r</mml:mi><mml:mi>u</mml:mi><
mml:mi>e</mml:mi><mml:mtext> </mml:mtext><mml:mi>p</mml:mi><mml:mi>o</mml:mi><
mml:mi>s</mml:mi><mml:mi>i</mml:mi><mml:mi>t</mml:mi><mml:mi>i</mml:mi><mml:mi>v
</mml:mi><mml:mi>e</mml:mi><mml:mi>s</mml:mi><mml:mtext> </mml:mtext></"#6D7376"
mml:mrow></mml:mfrac></mml:mrow>
</mml:math>
```

The answer…

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