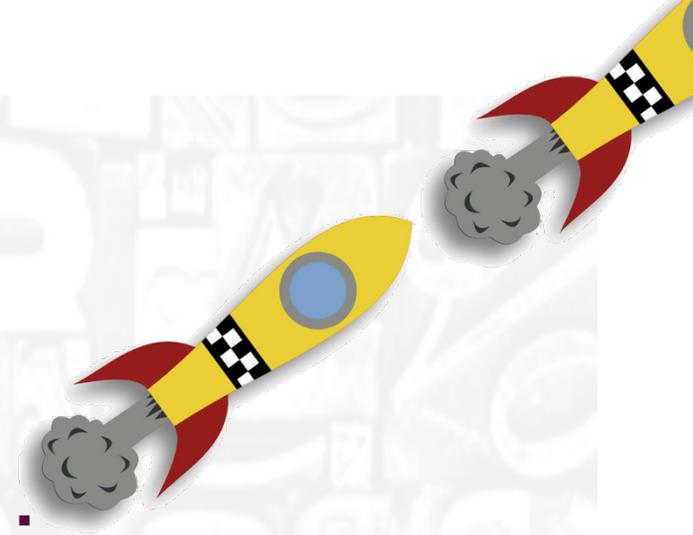


UK · CHINA · USA



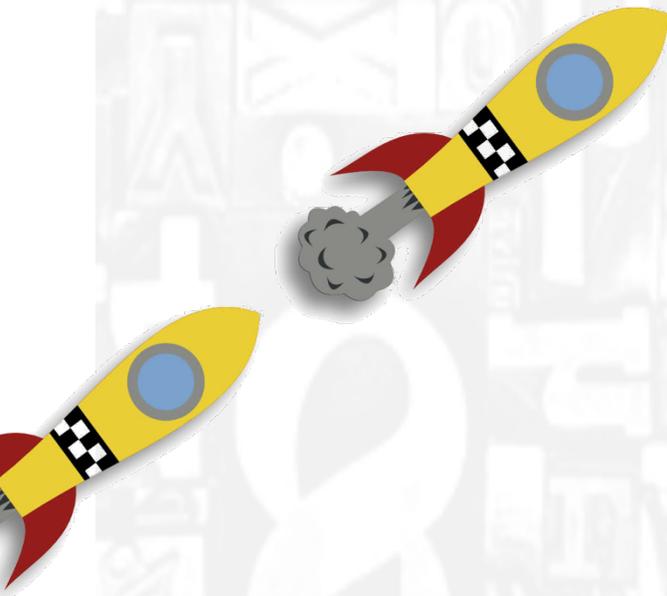
*The
Charlesworth
Group*



Partnering to Publish : XML: It's Not Rocket Science

(but *The more you know the more you know
you don't know*)

SSP/ARL Seminar
November 2010
Washington DC



Philippa J. Benson, Ph.D.
Director
Strategic & Business Development
The Charlesworth Group

UK · CHINA · USA



*The
Charlesworth
Group*

Who, me?



1984 – 1998

- MA and PH.D related to professional writing and research in literacy in science
- Head of Writing Center at Intercultural Communication Center at CMU
- Invited “Foreign Expert” in Technical Communication at key university in China (Shanghai Jiao Tong U.)
- Faculty at University of Pittsburgh overseeing WAC program, worked with librarians developing instructional materials

1998 - 2009

- Taught writing at Georgetown U., FAES, in mainland China (on publishing)
- Started information resources departments and launched publications at CI and TWS, both 501c3 organizations

2009 – Charlesworth, Director of Strategic and Business Development

UK · CHINA · USA



*The
Charlesworth
Group*

The Charlesworth Group



*The
Charlesworth
Group*

Family firm, founded in 1928.

Global Publishing Services

- **Journal & book typesetting (XML)**
- Marketing, rights, licensing, liaison (China)
- **Publishing services, language polishing**

Offices in

United Kingdom



United States



Mainland China



UK · CHINA · USA



*The
Charlesworth
Group*

Many thanks to....people who partner to solve problems

- **Steve Laverick, CWG**
- **Gurvinder Batra, Kiwi Technologies**
- **Dan O'Brien, American Chemical Society**
- **Devorah Bloom, Data Conversion Lab**
- **JATS-Con, NLM Journal Article Tag Suite Conference**



ACS
Chemistry for Life®

 **Data Conversion Lab**



KiwiTech

Think Global. Act Mobile.

<JATS-Con>

UK · CHINA · USA



*The
Charlesworth
Group*

Why are we here?

- Learn about existing partnerships
- **Better understand the role that you can play in providing publishing services for your organization**
- **Improve the quality of your publications through selective acquisition and editorial development.**
- Find out how to leverage resources within your institution
- Cost-effectively market and promote your published products

UK · CHINA · USA



The Charlesworth Group

Why back to basics?

Publarian?



Liblisher?



HTML

JATS

TAGS

SCEMA

XHTML

DTD

NLM

XML

XMP

RELAX NG

....blah, blah, blah, blah...

UK · CHINA · USA



The Charlesworth Group

XML allows your content to be created in one workflow, at one cost, for all readers

...in print



We can format your materials to be read....

... online



.... on the go



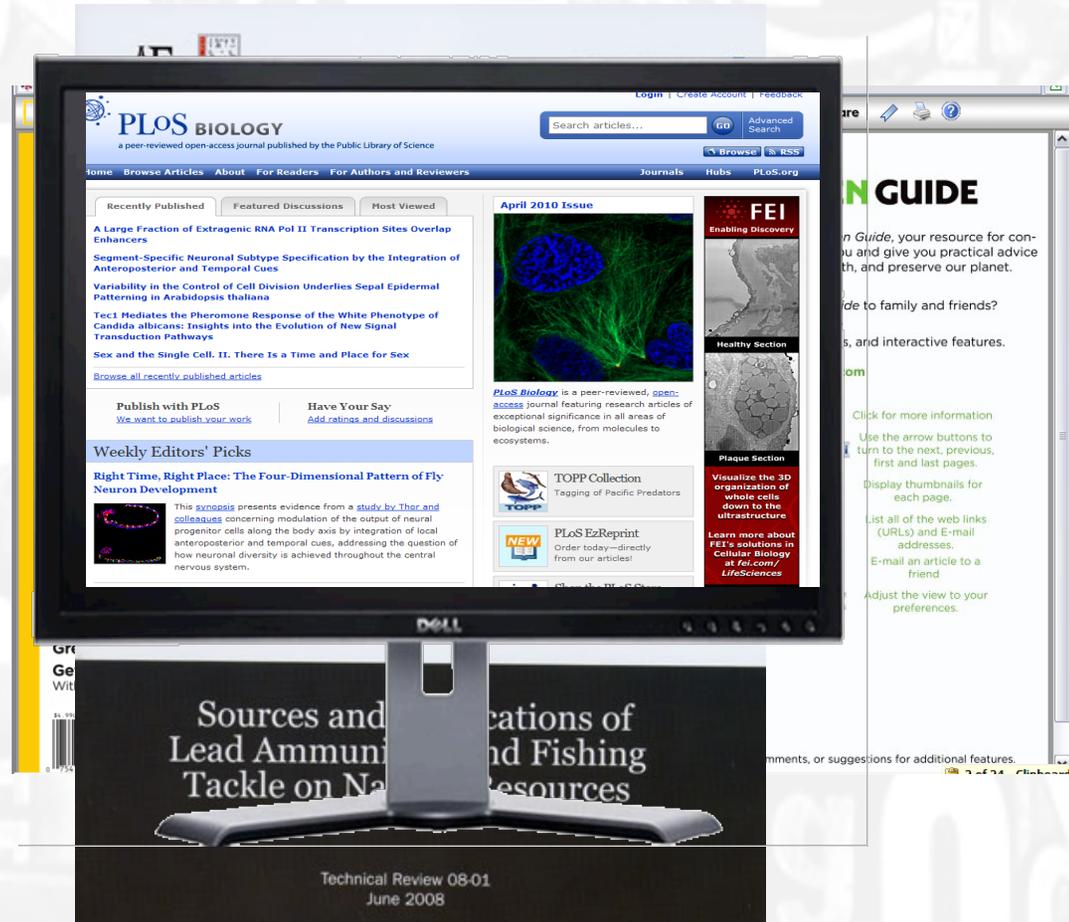
UK · CHINA · USA



The
Charlesworth
Group

Ancient Electronic Publishing

- PDF
- HTML
- Proprietary platforms



UK · CHINA · USA



The
Charlesworth
Group

HTML

- There is a heading.
- The heading is followed by two paragraphs.
- The 2nd paragraph has three line breaks.

<h1> Kubla Khan **</h1>**

<p> or a Vision in a Dream. A Fragment **</p>**

<p>In Xanadu did Kubla Khan a pleasure dome
decree **
**

Where Alf the sacred river ran **
**

Through caverns measureless to man **
**

Down to a sunless sea.**</p>**





XML tags are about content

```

<poem>
<poemtitle><firstname cent="13th id="p001">Kubla</
  firstname><honourarytitle rid="p001"> Khan</
  honourarytitle></poemtitle>
<author><surname>Coleridge</surname></author>
<stanza>
<line>In<placename type="mispronounced" location="North
  China" chinese="Shang-Tu">Xanadu</
  placename>did<firstname cent="13th"> Kubla</
  firstname><honourarytitle> Khan</honourarytitle> a
  pleasure dome decree</line>
<line>Where<river type="sacred" id="r001"
  location="China"> Alf</river> the sacred river
  ran</line>
<line>Through caverns measureless to man</
  line>
<line><direction rid="r001">Down</direction> to a
  sunless sea.</line></stanza>

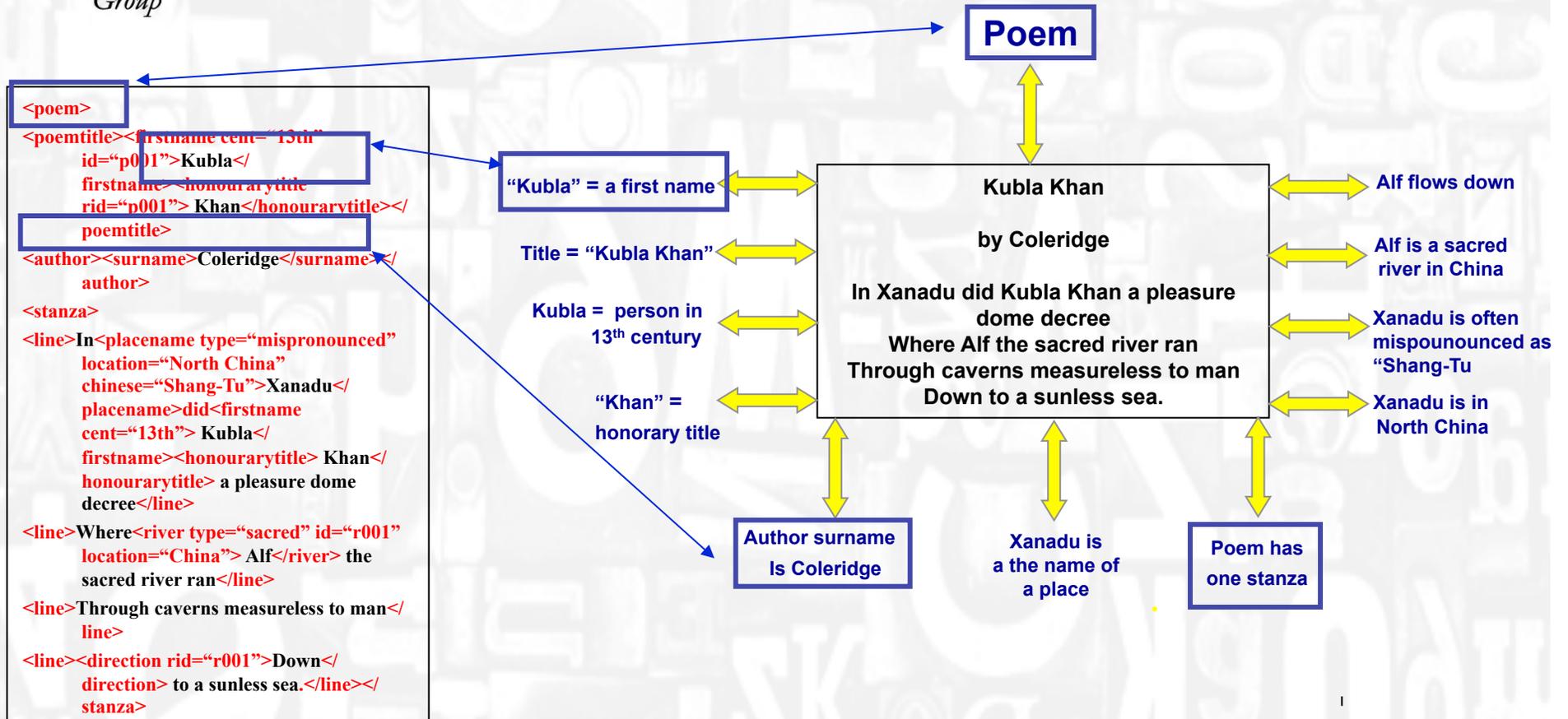
```

- This data is a poem
- The title of the poem is “Kubla Khan”
- Kubla is the first name of someone who lived in the 13th century
- Khan is an honorary title which applies directly to Kubla (and not to any other person mentioned, such as Coleridge)
- The surname of the poem’s author is Coleridge
- The poem has one stanza which consists of four lines
- Xanadu is the mispronounced name of a place in North China which the Chinese themselves call “Shang-Tu”
- Alf is the name of a river in China
- Alf is a sacred river
- The direction that Alf ran was “down”



The Charlesworth Group

XML tags are about content





The Charlesworth Group

One possible workflow



Document converted (.rtf)

Word (.doc) file is styled

INCREASED CANCER RISK IN HEAVY DRINKERS WITH THE ALCOHOL DEHYDROGENASE 3*1 ALLELE, POSSIBLY DUE TO SALIVARY ACETALDEHYDE

Short Title: ALCOHOL METABOLISM AND CANCER

Jukka-Pekka Vissapää¹, Karl Göter¹, Monica Bressova¹, Jianxi Li¹, Nils Homann², Christian Conrad³, Hiroko Inoue⁴, Matthias Tisch⁵, Karl Hörmann⁶, Sany Vakevainen⁷, Mihai Balazs⁸, Helmut K. Seitz¹

¹Laboratory of Alcohol Research, Liver Disease and Nutrition and Department of Medicine, Salem Medical Center, Heidelberg, Germany; ²Research Unit of Substance Abuse Medicine, University of Helsinki, Helsinki, Finland; ³ENT Hospital Mannheim, University of Heidelberg, Germany; ⁴Gastroenterology Unit, Department of Medicine, University of Lübeck, Germany; ⁵ENT Clinic, Federal Armed Forces Hospital, Ulm, Germany; ⁶Department of Biostatistics, University of Heidelberg, Heidelberg, Germany

Grant Support: This study was supported by a grant of the German-Chinese Medical Association and the Volkswagen Foundation awarded to Dr. Jianxi Li.

Abbreviations: ADH - alcohol dehydrogenase, UADTC - upper aerodigestive tract cancer, ALDH - aldehyde dehydrogenase, AA - acetaldehyde

ABSTRACT

Background: Chronic ethanol consumption is associated with an increased risk for upper aerodigestive tract cancer. Since acetaldehyde seems to be a carcinogenic factor associated with chronic alcohol

Abbrev-Title: gt

User-ID: g33449

DOI: na

Normal

Auxiliary Number: 2

ShortTitle: Interface between the intestinal environment and the nervous system

ShortAuthors: Lundgren

Banner: Supplement

Title: Interface between the intestinal environment and the nervous system

Abbreviation: Abbreviations: CCK, cholecystokinin; EC, enterochromaffin; cAMP, cyclic adenosine monophosphate; 5-HT, 5-hydroxytryptamine; CNS, central nervous system; NO, nitric oxide

Author: O Lundgren

Author: Professor O Lundgren, Department of Physiology, Sahlgrenska Academy, Göteborg

AbstractPara: Possible mechanisms by which the intestinal contents may influence afferent permeability of the intestinal epithelium

Para: Diffusion We cannot have references in the abstract, can you insert reference No 1 in the

Para: It is well established that the sensitivity of the muscle spindles in skeletal muscle is controlled from the central nervous system (CNS) via efferent fibres influencing the so-called intrafusal muscle cells. One can envisage a similar CNS control of the sensitivity of the endocrine cells of the gastrointestinal epithelium via receptors known to be present on endocrine cells. Activation of these receptors may change the sensitivity of endocrine cells to luminal stimuli by influencing their membrane potential.

Para: **Brush cells**

Para: There exists a specialised type of epithelial cell scattered within the epithelial lining of the respiratory and gastrointestinal tracts called the "brush cell", "tufted cell", or "caveolated cell". In the gastrointestinal tract, these cells have been found in the stomach and intestine and also in the excretory ducts of the gall bladder and pancreas. They are pear or bottle shaped with a broad germinular portion (fig 2). An apical microvillus tuft is in contact with the intestinal lumen. No granules of the type present in endocrine cells can be demonstrated in brush cells.[11]

Figure: Fig Name- g33449.F2-A

Figure: Figure 2 Schematic figure of a brush cell. The cell, located between enterocytes, has a bottle-like appearance. It lacks the granules at the basolateral part characteristic of endocrine cells

"Raw" Word Document

.rtf (styled)

Mains ▾ x

HeadA

HeadB

HeadC

HeadD

HeadE

List1

List2 List3

ListParaMore

Remark

FootNotes

EndNotes

Reference

Quote Para

ClearStyle

Title

Author

Front matter

Symbols

References

....and other

regularly

formatted

elements



The Charlesworth Group

One possible workflow



Document converted
(.rtf)



Document converted
(XML)



```

<poem><poemtitle><firstname cent="13th"
id="p001">Kubla</firstname><honourarytitle
rid="p001">Khan</honourarytitle></
poemtitle><author><surname>Coleridge</
surname></author><stanza><line>In<placename
type="mispronounced" location="North China"
chinese="Shang-Tu">Xanadu</placename>did
<firstname cent="13th">Kubla</firstname>
<honourarytitle> Khan</honourarytitle> a pleasure
dome decree</line><line>Where<river
type="sacred" id="r001" location="China"> Aif</
river> the sacred river ran</line><line>
Through caverns measureless to man</line> <line>
<direction rid="r001">Down</direction> to a sunless
sea.</line></stanza>
    
```

Styled (.rtf) file
Is converted to
XML



The Charlesworth Group

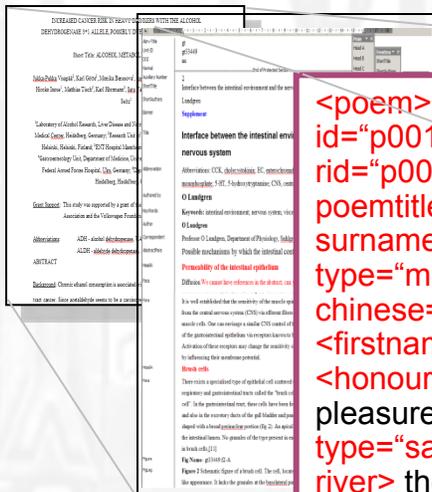
One possible workflow



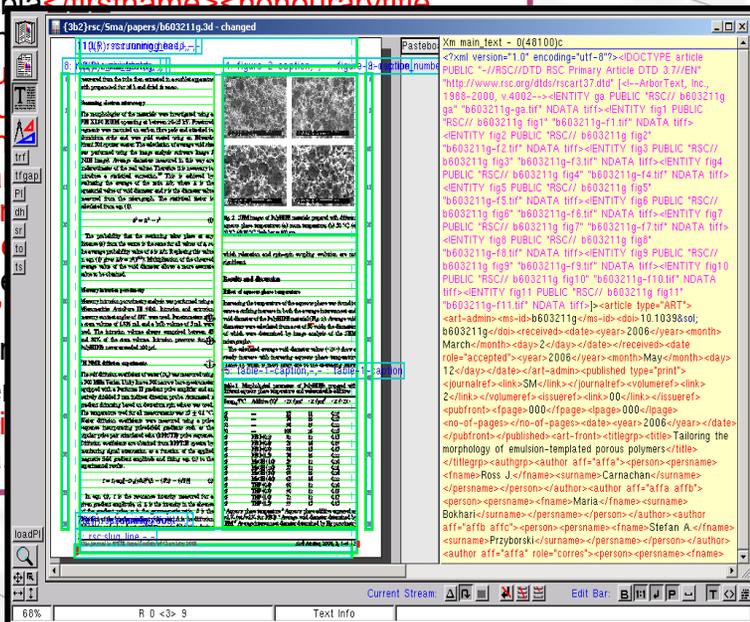
Document converted (.rtf)

Document converted (XML)

Customized output Print & web PDF, HTML



<poem><poemtitle><firstname cent="13th" id="p001">Kubla</firstname><honourarytitle><au surname></au type="mispronchinese="Sha <firstname cent="13th" id="p001">Kubla</firstname><honourarytitle pleasure dom type="sacred" river> the sac Through cave <line> <directi a sunless sea





The Charlesworth Group

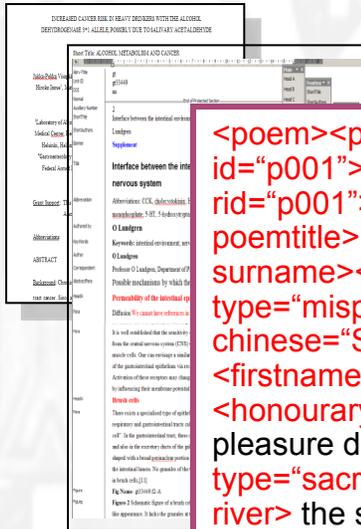
Another possible workflow



Document converted (.rtf)

Document converted (XML)

Customized output from XML: HTML, ePubs, apps



```

<poem><poemtitle><firstname cent="13th" id="p001">Kubla</firstname><honourarytitle rid="p001">Khan</honourarytitle></poemtitle><author><surname>Coleridge</surname></author><stanza><line>In<placename type="mispronounced" location="North China" chinese="Shang-Tu">Xanadu</placename>did<firstname cent="13th">Kubla</firstname><honourarytitle> Khan</honourarytitle> a pleasure dome decree</line><line>Where<river type="sacred" id="r001" location="China"> Alf</river> the sacred river ran</line><line>Through caverns measureless to man</line><line><direction rid="r001">Down</direction> to a sunless sea.</line></stanza>

```





XML tags are linked to DTDs

XML

- Is governed by **industry standards** and **rules** (e.g., NLM DTD), for archiving (e.g., PMC), and tagging (e.g., metadata)
- Allows **automatic inclusion** of standard STM electronic features such as DOIs.
- Allows for **full or selected automated services** including composition, design, and/or proofing
- Can be **driven by templates** or **tailored** to users to have full control of output

Acronyms of the Day

NLM = National
Library
of Medicine

**DTD = Data Tag
Definitions**

PMC = Pub Med
Central

DOI = Digital
Object
Identifier

UK · CHINA · USA



*The
Charlesworth
Group*

BUT.....

UK · CHINA · USA



*The
Charlesworth
Group*

The National Library of Medicine (NLM) DTD

Although the NLM DTD

is the gold standard in STM publishing

it still has multiple flavors

UK · CHINA · USA



*The
Charlesworth
Group*

Journal Article Tag Suite (JATS) makes up the NLM DTD

Archiving and Interchange Tag Set

Created to enable an archive to capture as many of the structural and semantic components of existing printed and tagged journal material as conveniently as possible, with no effort made to model any particular sequence or textual format

Journal Publishing Tag Set

Optimized for the archives that wish to regularize and control their content, not to accept the sequence and arrangement presented to them by any particular publisher

Article Authoring Tag Set

Designed for authoring new journal articles, where regularization and control of content is important

NCBI Book Tag Set

Written specifically to describe volumes for the NCBI online libraries

UK · CHINA · USA



*The
Charlesworth
Group*

The DTD spectrum



Archive intellectual
content or Article file??

The Journal Publishing set is too restrictive
for some, and not prescriptive enough to
be a good Authoring model.



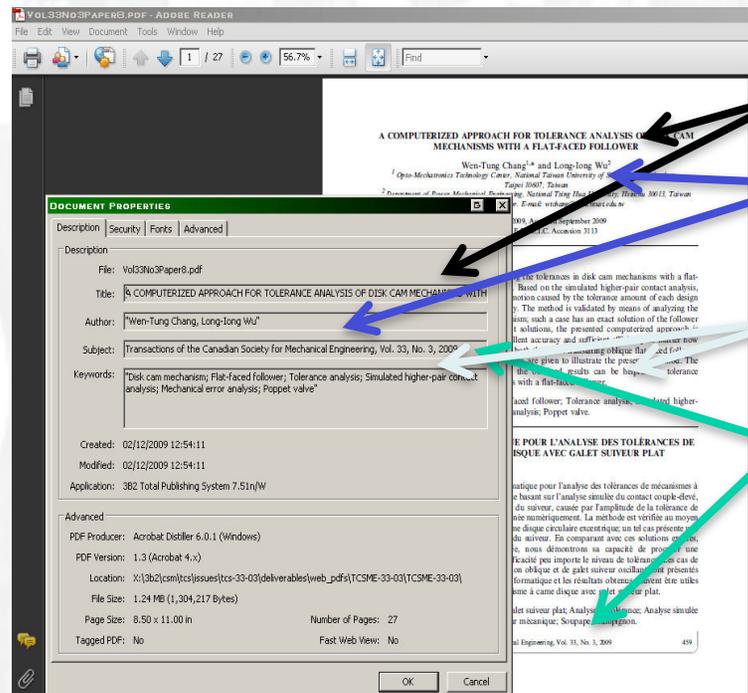
ACS: Tags, tags, and more tags

As-Is	Extended	Reduced	Customized	Built From	Informed By
Public version is used without changes or modification	Superset of public tagset is used	Subset of the public schema is used	Combo of Extensions + Reductions	Substantial changes: renamed tags, altered hierarchies, etc	Only the design philosophy of public tagset is used
<code><xyz></code> <code><a/></code> <code></code> <code></xyz></code>	<code><xyz></code> <code><a/></code> <code></code> <code><c/></code> <code></xyz></code>	<code><xyz></code> <code><a/></code> <code></code> <code></xyz></code>	<code><xyz></code> <code><a/></code> <code></code> <code><c/></code> <code></xyz></code>	<code><abc></code> <code><a></code> <code></code> <code></code> <code></abc></code>	<code><abc></code> <code><aa/></code> <code><bb/></code> <code><cc/></code> <code></abc></code>
XML is compatible	Public → Custom	Public ← Custom	XML not compatible?	XML not compatible!	XML not compatible!



XMP: eXtensible Metadata Platform

- Supports embedding of metadata in PDF files
- Inexpensive
- Human & machine readable
- Futureproofs PDFs for use in new web applications



Article title

Authors

Keywords

Article/issue info

UK · CHINA · USA



*The
Charlesworth
Group*

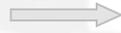
ePub and XHTML

- **ePub is based on XHTML**
- **XHTML is a combination of HTML and XML and includes all the elements in HTML with base syntax from XML.**
- **ePub for Sony, Apple, and other readers all slightly different**
- **ePub varieties (XHTML) differ primarily by metadata used**

UK · CHINA · USA



The Charlesworth Group



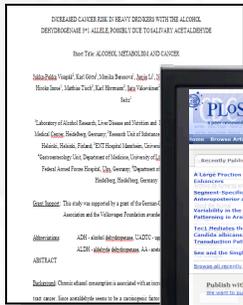
Document converted To HTML



HTML reflowed into InDesign (PDF created)



Content extracted from InDesign, converted to XML deposit to PubMed



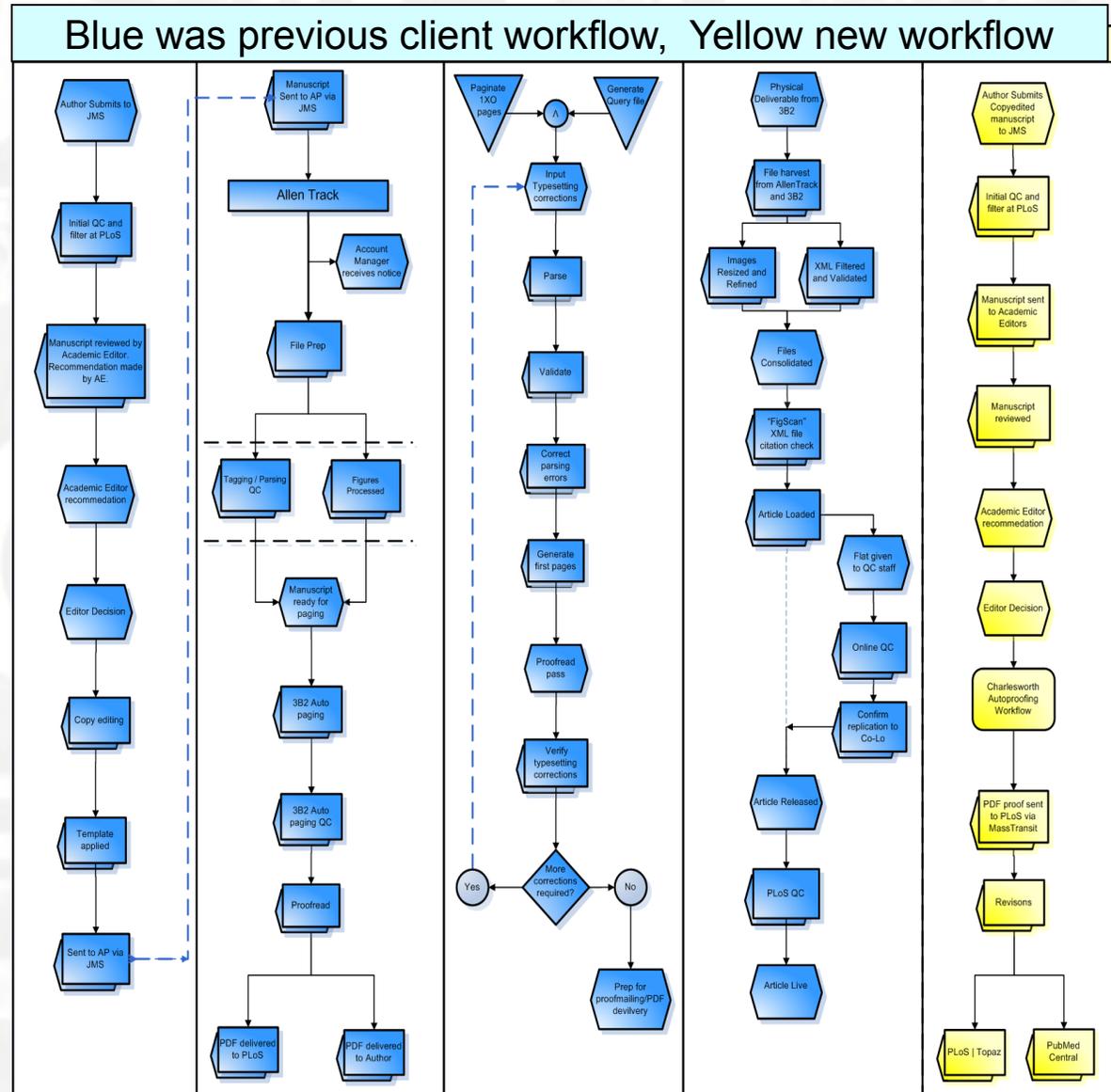
<article><poemtitle><firstname cent="13th" id="p001">Kubla</firstname><honourarytitle rid="p001">Khan</honourarytitle></poemtitle><author><surname>Coleridge</surname></author><stanza><line>In<placename type="mispronounced" location="North China" chinese="Shang-Tu">Xanadu</placename> did <firstname cent="13th">Kubla </firstname> <honourarytitle> Khan </honourarytitle> a pleasure dome decree </line><line>Where <river type="sacred" id="r001" location="China"> Alf</river> the sacred river ran</line><line>Through caverns measureless to man</line> <line> <direction rid="r001"> Down</direction> to a sunless sea .</line></ stanza>



The Charlesworth Group

A revamped workflow

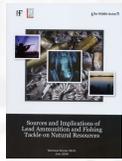
- ✓ **Workflow is streamlined and condensed.**
- ✓ **A typical fast track schedule is 4 - 6 days (2 days to 1XO, 2 days for revisions if required, 2 days to delivery of online deliverables.)**
- ✓ **Other journals typically take 23 – 28 days from initial submission to online Release**





The Charlesworth Group

The tagging spectrum



```

<poem><poemtitle><firstna
me cent="13th id="p001">
Ku
fir
rid
ho
po
<s
on
au
</poemtitle><firstna
me cent="13th id="p001">
Kubla</
firstname><honourytitle
rid="p001"> Khan </
honourytitle></
poemtitle><author>
<surname> Coleridge</
surname></
author><stanza>

```

Untagged Content

PDF

HTML

XMP
PDF

XHTML

XML
(Minus) (Plus)

Rich tagged content

Format

Content



UK · CHINA · USA



*The
Charlesworth
Group*

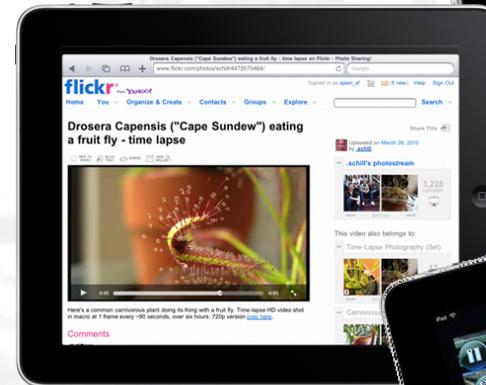
Readers are beginning to expect more of their texts



To communicate



To watch



To interact





So what's XML got to do with it?

- **DTDs / XML comes in many flavors**
- **Pace of development is mind-blowingly fast**
- **What readers want is changing quickly**
- **Partners can help you keep pace, make informed decisions, reduce costs, streamline workflows, and advocate**
- **Partners can create content in target formats and functionalities**

Questions?



pbenson@charlesworth.com



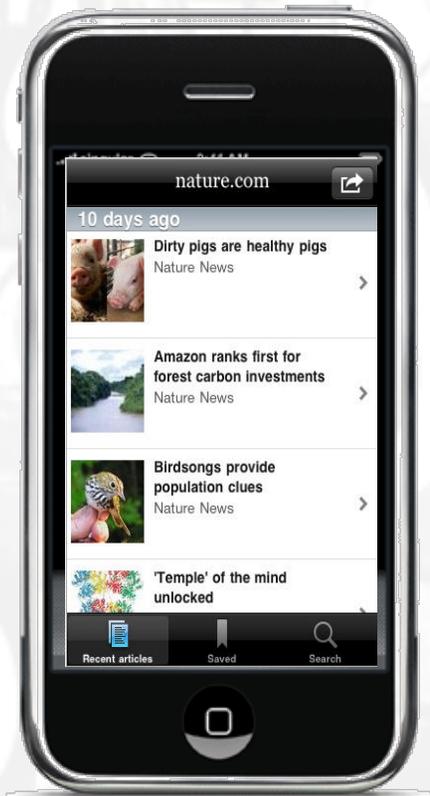
*The
Charlesworth
Group*

UK · CHINA · USA



*The
Charlesworth
Group*

Some apps are glorified browsers



UK · CHINA · USA



*The
Charlesworth
Group*

Some apps are self-contained books

