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
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
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
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
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
Why back to basics?

Publarian?

Liblisher?

HTML

DTD

RELAX NG

JATS

TAGS

SCEMA

NLM

XML

XHTML

XMP

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

We can format your materials to be read:

- ...in print
- ...online
- .... on the go
Ancient Electronic Publishing

- PDF
- HTML
- Proprietary platforms
<h1>Kubla Khan</h1>

or a Vision in a Dream. A Fragment

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.
<poem>

In <placename type="mispronounced" location="North China" chinese="Shang-Tu">Xanadu</placename> did <firstname cent="13th" id="p001">Kubla</firstname><honourarytitle rid="p001">Khan</honourarytitle> a pleasure dome decree</line>

Where <river type="sacred" id="r001" location="China">Alf</river> the sacred river ran</line>

Through caverns measureless to man</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 honourary 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”
Kubla Khan
by Coleridge

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.
One possible workflow

“Raw” Word Document

Word (.doc) file is styled

Title
Author
Front matter
Symbols
References
….and other regularly formatted elements

Document converted (.rtf)

.rtf (styled)
In Xanadu did Kubla Khan a pleasure dome decree
Where the sacred river ran
Through caverns measureless to man
Down to a sunless sea.

Kubla Khan

Xanadu

Alf

China

North China

Shang-Tu

Xanadu
<poem>
<poemtitle><firstname cent="13th" id="p001">Kubla</firstname><honourarytitle id="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 <direction rid="r001">Down</direction> to a sunless sea.
</line></stanza>

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 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**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLM</td>
<td>National Library of Medicine</td>
</tr>
<tr>
<td>DTD</td>
<td>Data Tag Definitions</td>
</tr>
<tr>
<td>PMC</td>
<td>Pub Med Central</td>
</tr>
<tr>
<td>DOI</td>
<td>Digital Object Identifier</td>
</tr>
</tbody>
</table>
BUT.....
Although the NLM DTD is the gold standard in STM publishing, it still has multiple flavors.
Journal Article Tag Suite (JATS) makes up the NLM DTD

<table>
<thead>
<tr>
<th>Tag Set</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archiving and Interchange Tag Set</td>
<td>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</td>
</tr>
<tr>
<td>Journal Publishing Tag Set</td>
<td>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</td>
</tr>
<tr>
<td>Article Authoring Tag Set</td>
<td>Designed for authoring new journal articles, where regularization and control of content is important</td>
</tr>
<tr>
<td>NCBI Book Tag Set</td>
<td>Written specifically to describe volumes for the NCBI online libraries</td>
</tr>
</tbody>
</table>
The DTD spectrum

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

Archive intellectual content or Article file??

The DTD spectrum
ACS: A mental model
## ACS: Tags, tags, and more tags

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

XML is compatible
• Supports embedding of metadata in PDF files
• Inexpensive
• Human & machine readable
• Futureproofs PDFs for use in new web applications
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
Kubla Khan

In Xanadu did a pleasure dome decree

Where the sacred river ran

Through caverns measureless to man
down to a sunless sea.
In Xanadu did Kubla Khan a pleasure dome decree
Where Al the sacred river ran
Through caverns measureless to man
Down to a sunless sea.
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.
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
Some apps are glorified browsers
Some apps are self-contained books