



Single source, multiple outputs: how XML will transform publishing (again)

STM E-Production
London, December 3, 2009

Kevin Cohn
Director of Client Services

Atypon is a leading provider of software used by publishers to manage, deliver, and monetize their digital content.

Customers include the American Chemical Society, Annual Reviews, the Institute of Electrical and Electronics Engineers, and the *New England Journal of Medicine*.

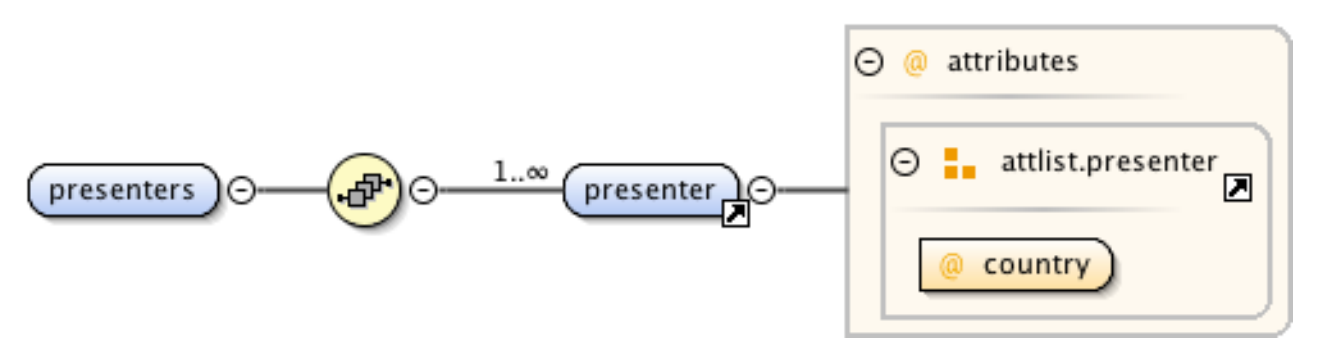


All things X—a primer

XSD

“[XML Schema] provide a means for defining the structure, content, and semantics of XML documents.”

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">
  <xs:element name="presenters">
    <xs:complexType>
      <xs:sequence>
        <xs:element maxOccurs="unbounded" ref="presenter"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="presenter">
    <xs:complexType mixed="true">
      <xs:attributeGroup ref="attlist.presenter"/>
    </xs:complexType>
  </xs:element>
  <xs:attributeGroup name="attlist.presenter">
    <xs:attribute name="country" use="required"/>
  </xs:attributeGroup>
</xs:schema>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT presenters (presenter+)>
<!ELEMENT presenter (#PCDATA)>
<!ATTLIST presenter
    country CDATA #REQUIRED
>
```


XML

You know this.

```
<?xml version="1.0" encoding="UTF-8"?>
<presenters xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="file:/example.xsd">
  <presenter country="us">Kevin Cohn</presenter>
  <presenter country="uk">Aviva Weinstein</presenter>
</presenters>
```

XPath

“A language for addressing parts of an XML document.”

```
//presenter
```

```
Kevin Cohn  
Aviva Weinstein
```

```
//presenter[@country='us']
```

```
Kevin Cohn
```

```
//presenter[@country='us' or @country='uk']
```

```
Kevin Cohn  
Aviva Weinstein
```


XSLT

“[XSL Transformations] is a language for transforming XML documents into other XML documents.”

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
  <xsl:output omit-xml-declaration="yes" />
  <xsl:template match="presenters">
    <h1>Presenters</h1>
    <xsl:for-each select="//@country">
      <xsl:call-template name="presenter-list">
        <xsl:with-param name="country" select="."></xsl:with-param>
      </xsl:call-template>
    </xsl:for-each>
  </xsl:template>
  <xsl:template name="presenter-list">
    <xsl:variable name="country" select="." />
    <h2><xsl:value-of select="$country" /></h2>
    <ul>
      <xsl:for-each select="//presenter[@country=$country]">
        <li><xsl:value-of select="text()" /></li>
      </xsl:for-each>
    </ul>
  </xsl:template>
</xsl:stylesheet>
```

```
<h1>Presenters</h1>
<h2>us</h2>
<ul>
  <li>Kevin Cohn</li>
</ul>
<h2>uk</h2>
<ul>
  <li>Aviva Weinstein</li>
</ul>
```


Presenters

us

- Kevin Cohn

uk

- Aviva Weinstein

XML is more than markup

Although XML workflows are increasingly prevalent in STM publishing, many publishers are not taking full advantage of the opportunities they afford.

In this presentation, the speaker uses his early experiences in moving software documentation to an XML workflow as an example of what can be tried in STM publishing.

The advantages and challenges of this project are discussed.

Current state

- Authoring in FrameMaker (proprietary format; don't try exporting XML)
- With few exceptions, each document exists as a single FrameMaker file
- We generate PDFs; there's no HTML output or embedded documentation
- Documentation doesn't account for varying degrees of user permissions, etc.

Desired state (advantages)

- Authoring in </oXygen/> (DocBook XSD)
- Modularized content (XInclude)
- PDFs and HTML (DocBook XSLs)
- Embedded documentation (custom XSLs)
- Account for variations in application state (XSL parameters)

Challenges

- Modularization can be tricky, and after a while there are diminishing returns
- Deciding what tagging to use for state variations is difficult
- Technical writers need to become content management and modeling experts
- We're probably trying to achieve too many things at the same time

Author Services User Guide

Version 1027, November 5, 2009

Copyright © 2009 Atypion Systems, Inc.

Abstract

This document describes Literatum's Author Services feature set. The intended audience is publisher administrators. The document assumes that the reader has working knowledge of XML, the Literatum Admin Tool (including the file upload and reports interfaces), and Literatum identities, access tokens, and licenses.

Table of Contents

1. Author Identification	1
1.1. Identifying Authors In Content Submissions	1
1.2. Identifying Authors In File Uploads	2
1.3. Matching Authors to Person Identities	2
1.4. The <i>All Authors</i> Group Identity	2
2. Publication Notification and Eprints	3
2.1. Eprint Configuration Options	3
3. The Author Services Panel	3
4. Reports	4

1. Author Identification

This section describes how authors are identified in content submissions and file uploads, and matched to person identities. Authors must be identified and matched to a person identity before they can make use of the author services functionality.

1.1. Identifying Authors In Content Submissions

Authors are identified in article, book, and chapter content submissions as follows:

```
//contrib[@contrib-type='author']
```

For each author, the following information is extracted:

- *Prefix*: name/prefix
- *Given names*: name/given-names
- *Surname*: name/surname
- *Suffix*: name/suffix
- *Degrees*: degrees
- *Email address*: email
- *Language preference*: `//*[@xml:lang`

STM publishers could
be trying this, too.

STM adoption

STM adoption

- XML as source format: **Widely adopted**

STM adoption

- XML as source format: Widely adopted
- Modularizing content: N/A (yet)

STM adoption

- XML as source format: **Widely adopted**
- Modularizing content: **N/A (yet)**
- Different output formats: **Widely adopted**

STM adoption

- XML as source format: **Widely adopted**
- Modularizing content: **N/A (yet)**
- Different output formats: **Widely adopted**
- Delivering content differently depending on context or user preference: **Room for improvement**

ASCII	A “plain text” abstract, i.e., without special characters or equations, so the abstract can be sent in email or displayed on primitive browsers
executive-summary	A non-technical summation of the major findings of the article
graphical	A pictorial representation such as a picture or a video
editor	An abstract written by an editor, not an author
key-points	An abstract which lists the key points made by the article
objectives	An abstract used for Learning Objectives or article objectives
section	An abstract containing the titles of an article’s sections; following each title, that section is summarized.
short	An abbreviated form of the abstract, for example, for use inside a generated Table of Contents, or to be returned in addition to the article title during a search
stereochemical	An abstract containing only the details of a chemical compound, for example, a “stereochem” abstract
summary	Summation of the article, typically used in conjunction with other types of abstracts
teaser	A short abstract specifically written to create interest in the reader
toc	A very short abstract, usually only a line or two long, that is displayed in a Table of Contents
web-summary	Short summary intended for distribution on a website

Identification of uncommon targets for inhibition, such as virulence factors, represents an emerging approach for combating the problem of antibiotic resistance among bacteria. Unfortunately, the lack of effective systems for the discovery and evaluation of inhibitors for such targets has considerably slowed progress. A recent article in *ACS Chemical Biology*, however, details the development of an *in vivo* based high-throughput screening strategy for identification of small molecule inhibitors of wall teichoic acid biosynthesis.

Targeting Wall Techoic Acid Biosynthesis: An *in Vivo* Based High-Throughput Screen for Small Molecule Inhibitors

Wenlan Chen, Robert Woodward and Peng George Wang

pp 893-894

Publication Date (Web): November 4, 2009 (Points of View)

DOI: 10.1021/cb900259w

Abstract

 Full Text HTML

 Hi-Res PDF [666K]

 PDF w/ Links [162K]

 Sponsored Access

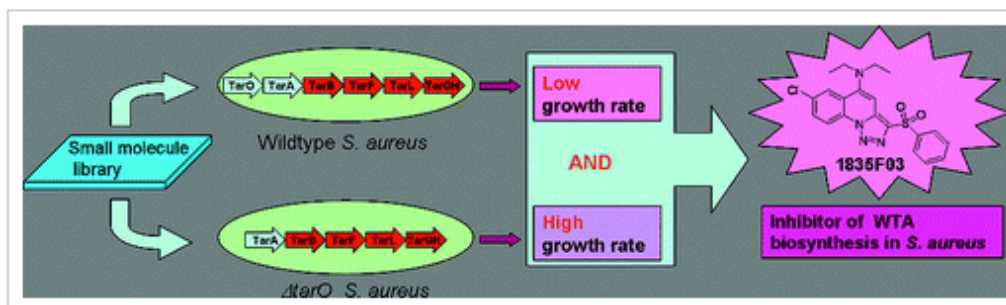


Figure 1 of 3

 Next

Hide Thumbnails 

ABSTRACT

Natural products containing carbon-phosphorus bonds (phosphonic and phosphinic acids) have found widespread use in medicine and agriculture. Recent years have seen a renewed interest in the biochemistry and biology of these compounds with the cloning of the biosynthetic gene clusters for several family members. This review discusses the commonalities and differences in the molecular logic that lie behind the biosynthesis of these compounds. The current knowledge regarding the metabolic pathways and enzymes involved in the production of a number of natural products, including the approved antibiotic fosfomycin, the widely used herbicide phosphinothricin (PT), and the clinical candidate for treatment of malaria FR-900098, is presented. Many of the enzymes involved in the biosynthesis of these compounds catalyze chemically and biologically unprecedented transformations, and a wealth of new biochemistry has been revealed through their study. These investigations have also suggested new strategies for natural product discovery.

SUMMARY POINTS

1. Phosphonates and phosphinates function by mimicking phosphate esters or anhydrides or carboxylate groups in enzyme substrates. As such, a large number of enzymes can be potential targets of this class of compounds.
2. The number of unprecedented reactions involved in the biosynthesis of fosfomycin, phosphinothricin, and FR-900098 is an indication of the wealth of novel biochemistry used in the biosynthesis of this class of compounds.
3. Phosphoenolpyruvate (PEP) mutase catalyzes the C-P bond-forming step in all naturally occurring phosphonates for which the gene clusters are currently known. Hence, degenerate primers for PEP mutase can be used for the discovery of new phosphonate-encoding gene clusters and new natural products.
4. Given the current commercial use of phosphonates and phosphinates in medicine and agriculture, discovery of new naturally occurring compounds beyond the 20 or so currently known structures may provide an important untapped source of new products for human use.

More possibilities

- Medical content can be displayed differently to practitioners and researchers
- Reference entries can have basic (undergrad) and advanced (postgrad) “views”
- An entire publication’s text can be reordered based on user preference

Recommendations

- Define the product before you invest in tagging and development
- Develop different kinds of teasers (abstracts) for different paths to content
- Start slow with a selection of articles, a single journal, and so on

Remember: XML is more than markup; it's a powerful set of languages that allows you to deliver your content to readers in a way that maximizes its relevancy and value.



Thank you

Kevin Cohn
kcohn@atypon.com