

Welcome to the 17th Annual  
**EMUG BOSTON**

#EMUG2019



# EMUG 2019

## XML Bootcamp

**Charles O'Connor**  
**Business Systems Analyst**  
**[coconnor@ariessys.com](mailto:coconnor@ariessys.com)**

# What is XML?

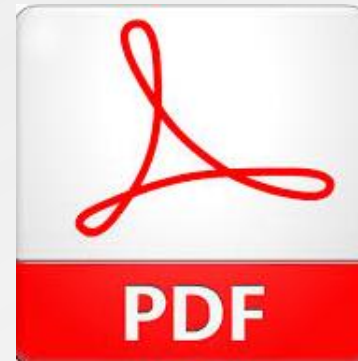
- XML is annotated text

# XML is not

- A proprietary binary format like...



or



# XML is a Markup Language

- Like HTML

```
<h1>This is a first-level heading</h1>
```

```
<h2>This is a second level heading</h2>
```

```
<p>This is a paragraph of text.</p>
```

- or Markdown

```
#This is a first-level heading
```

```
##This is a second-level heading
```

```
This is a paragraph of text.
```

# XML Contains

- Elements
  - What is it?

```
Please see our companion paper,  
<related-article>Jones et al,  
2017</related-article>.
```

... and, of course, text

- Attributes
  - What kind? Which one? Linked where?

```
<related-article id="ra1" related-article-  
type="companion" ext-link-type="doi"  
xlink:href="10.1175/pone.547862">Jones et  
al, 2017</related-article>
```

# XML Links

- Link

```
<xref ref-type="fig" rid="f1">Fig. 1A-1C</xref>
```

- Target

```
<fig id="f1"><label>Figure 1</label>  
<caption><p>This figure represents the  
universe.</p></caption>  
<graphic xlink:href="pone.25648512_f1"/></fig>
```

# XML vs HTML

- HTML Describes How Text Should Be Displayed

```
<h1>The Biosemiotics of Cheese Production</h1>
<ul>
  <li>Tom Jones</li>
  <li>Neil Sedaka</li>
</ul>
```

- XML Describes Meaning

```
<article-title>The Biosemiotics of Cheese Production</article-title>
<contrib>
  <given-names>Tom</given-names><surname>Jones</surname>
</contrib>
<contrib>
  <given-names>Neil</given-names><surname>Sedaka</surname>
</contrib>
```



# XML vs HTML

- Database Link in HTML

```
<a href="https://www.ncbi.nlm.nih.gov/nuccore/NM_167141.3">
```

- Database Link in XML

```
<ext-link ext-link-type="gen" xlink:href="NM_167141.3">
```

# Content → Presentation

- XML Content

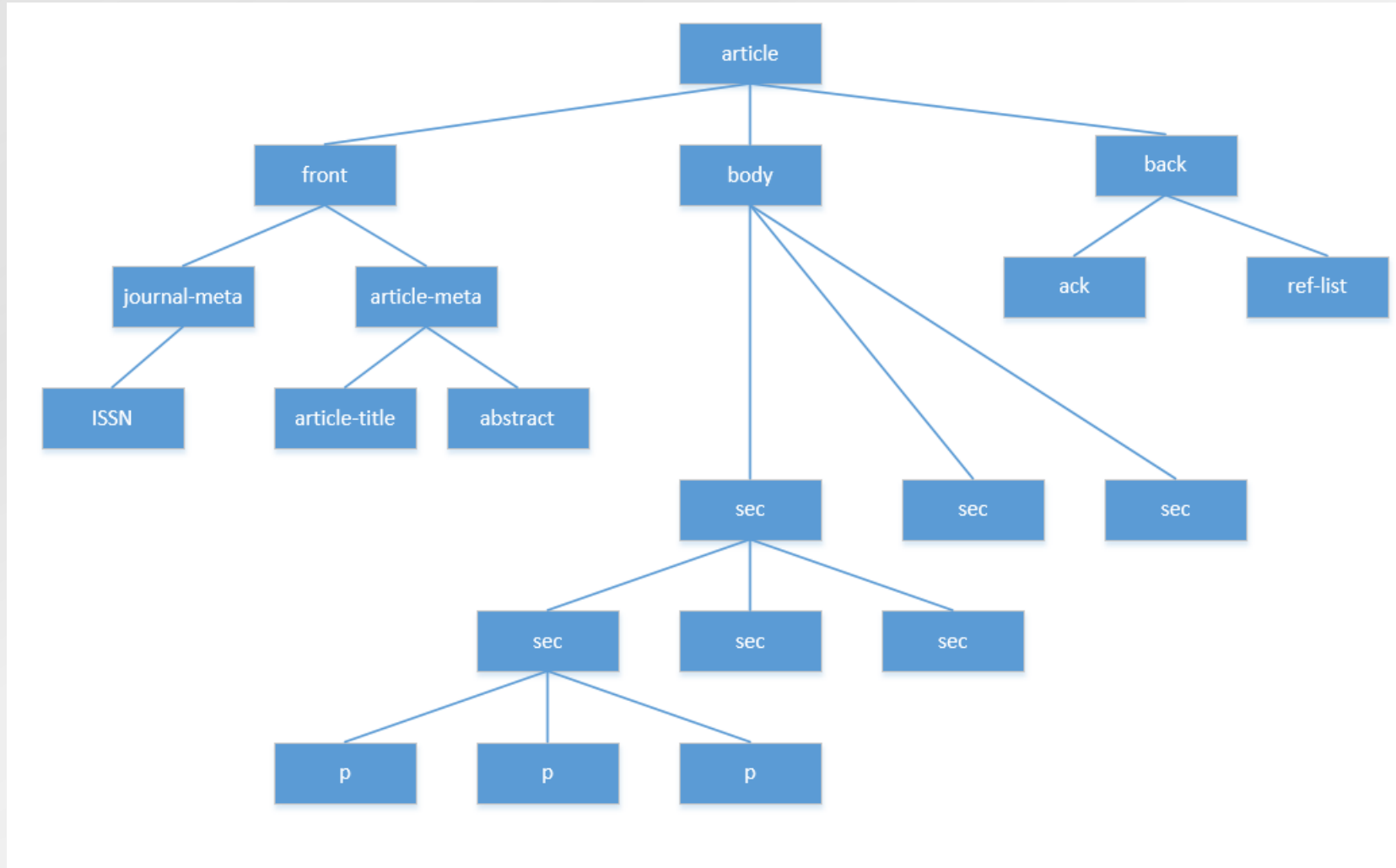
```
<history>
  <date date-type="received">
    <day>22</day><month>12</month><year>2016</year>
  </date>
  <date date-type="accept">
    <day>21</day><month>04</month><year>2017</year>
  </date>
</history>

<date date-type="pub" publication-format="online">
  <day>22</day><month>05</month><year>2017</year>
</date>
```

- Presentation Online

<b>Received:</b> December 22, 2016; <b>Accepted:</b> April 21, 2017; <b>Published:</b> May 22, 2017
---

# XML is a Tree



# A DTD Defines the XML

- DTD is short for Document Type Definition
- XML – eXtensible Markup Language
- What elements and attributes can appear in a document
- What is the order of elements
- What can appear in elements
  - Only other elements?
  - Only text?
  - Text and other elements?
- DTDs include JATS, DocBook, DITA, etc.

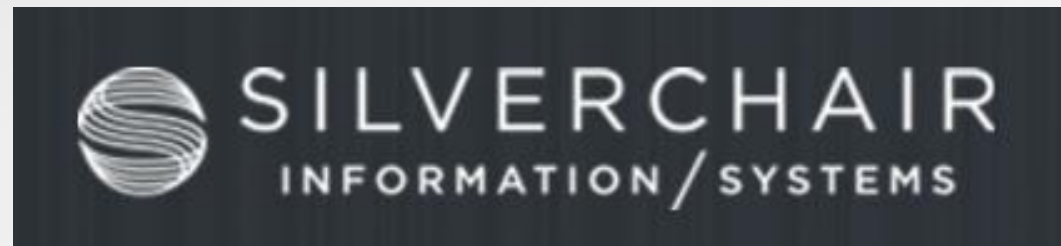
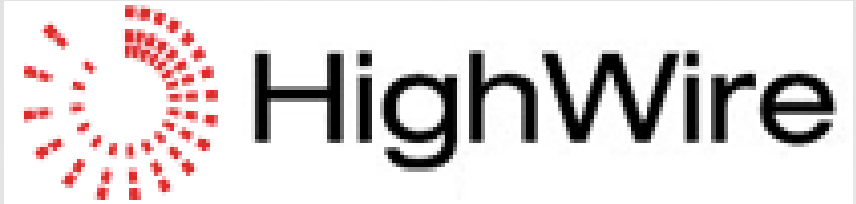
# The JATS DTD

- Developed by the U.S. National Library of Medicine
- First version released in March 2003
- The Journal Article Tag Set became a NISO standard in July 2012
- JATS is the de facto standard for journal articles in scholarly publishing—not just science, technology, and medicine

# Applications of JATS

- Metadata Transfer out of EM: Aries uses JATS XML as an exchange medium to transmit metadata from EM to customer systems, preprint servers, and vendors.
- Submission Import into EM: Aries uses JATS XML to import submission metadata from Submission Partners, preprint servers, and other peer review systems.
- Submission Import into ProduXion Manager: Aries uses JATS XML to import submissions from a peer review system directly into PM.
- MECA (Manuscript Exchange Common Approach): Aries supports the import and export of MECA packages, which include a JATS XML file.
- Archiving: Portico stores journal articles in JATS to preserve them after journals cease publication.
- Online Hosting: JATS XML is the primary vehicle for content delivery to online hosts.
- Layout: XML can be used to drive the production of composed pages.

# Online Hosts



# Metadata Initiatives



```
<contrib-id contrib-id-type="orcid"  
authenticated="true">https://orcid.org/0000-0002-6046-2077</contrib-id>
```



# Metadata Initiatives

The logo for CHORUS, featuring the word "CHORUS" in a blue, sans-serif font. The letter "O" is stylized with a blue arrow pointing upwards and to the right.

```
<funding-group specific-use="FundRef">  
  <award-group>  
    <funding-source>  
      <institution-wrap>  
        <institution>U.S. Department of Energy</institution>  
        <institution-id>https://dx.doi.org/10.13039/100000015</institution-id>  
      </institution-wrap>  
    </funding-source>  
    <award-id>DE-FC26-07NT43098</award-id>  
  </award-group>  
</funding-group>
```

# Metadata Initiatives

The logo for CHORUS, featuring the word "CHORUS" in a blue, sans-serif font. The letter "O" is stylized with a blue arrow pointing upwards and to the right.

```
<license>  
<ali:license_ref xmlns:ali="https://www.niso.org/schemas/ali/1.0/"  
specific-use="am" start_date="2015-07-01">  
https://creativecommons.org/licenses/by/4.0/</ali:license_ref>  
</license>
```

# Production Workflows

## Diagnosis of Cystic Fibrosis: Consensus Cystic Fibrosis Foundation

Philip M. Farrell, MD, PhD<sup>1</sup>, Terry B. White, PhD<sup>2</sup>, Clement L. Ren, MD  
Nico Derichs, MD<sup>3</sup>, Michelle Howenstine, MD<sup>3</sup>, Susanna /  
Margaret Rosenfeld, MD, MPH<sup>4</sup>, Isabelle Sermet-Gaudelus, M  
Bruce C. Marshall, MD<sup>2</sup>, and Patrick

**Objective** Cystic fibrosis (CF), caused by mutations in the CF gene, continues to present diagnostic challenges. Newborn screening has prompted a reconsideration of the diagnosis criteria.

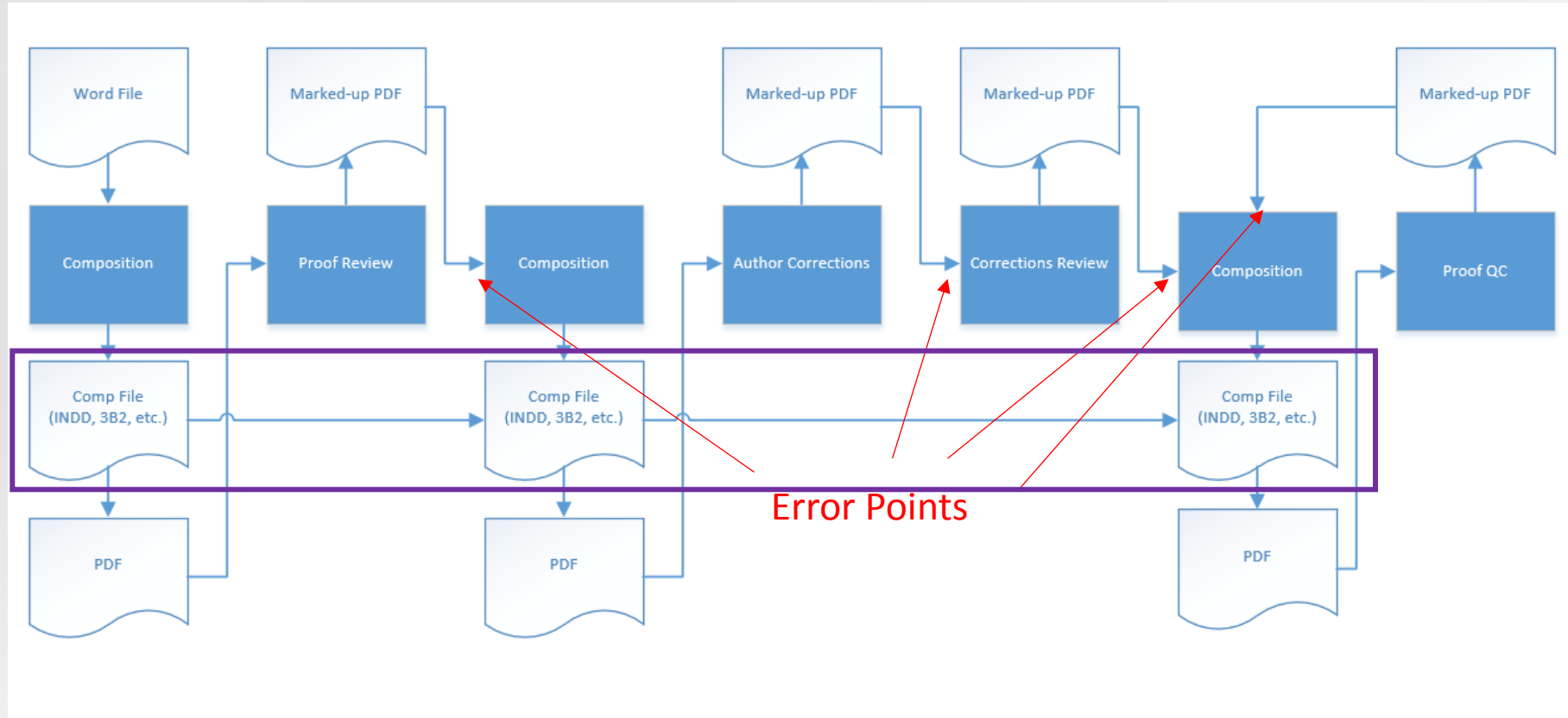
**Study design** To improve diagnosis and achieve standardized care, a committee of 32 experts in CF diagnosis from 9 countries met to discuss the diagnosis of CF and to clarify diagnostic criteria and CFTR mutations. An a priori threshold of ≥80% affirmative votes was required for a statement.

**Results** After reviewing relevant literature, the committee convened a conference, consensus statements were developed by an executive committee and approved 27 of 28 statements, 7 of which need to be established by evaluation of CFTR function with a variant identified in the Clinical and Functional Translation of CFTR.

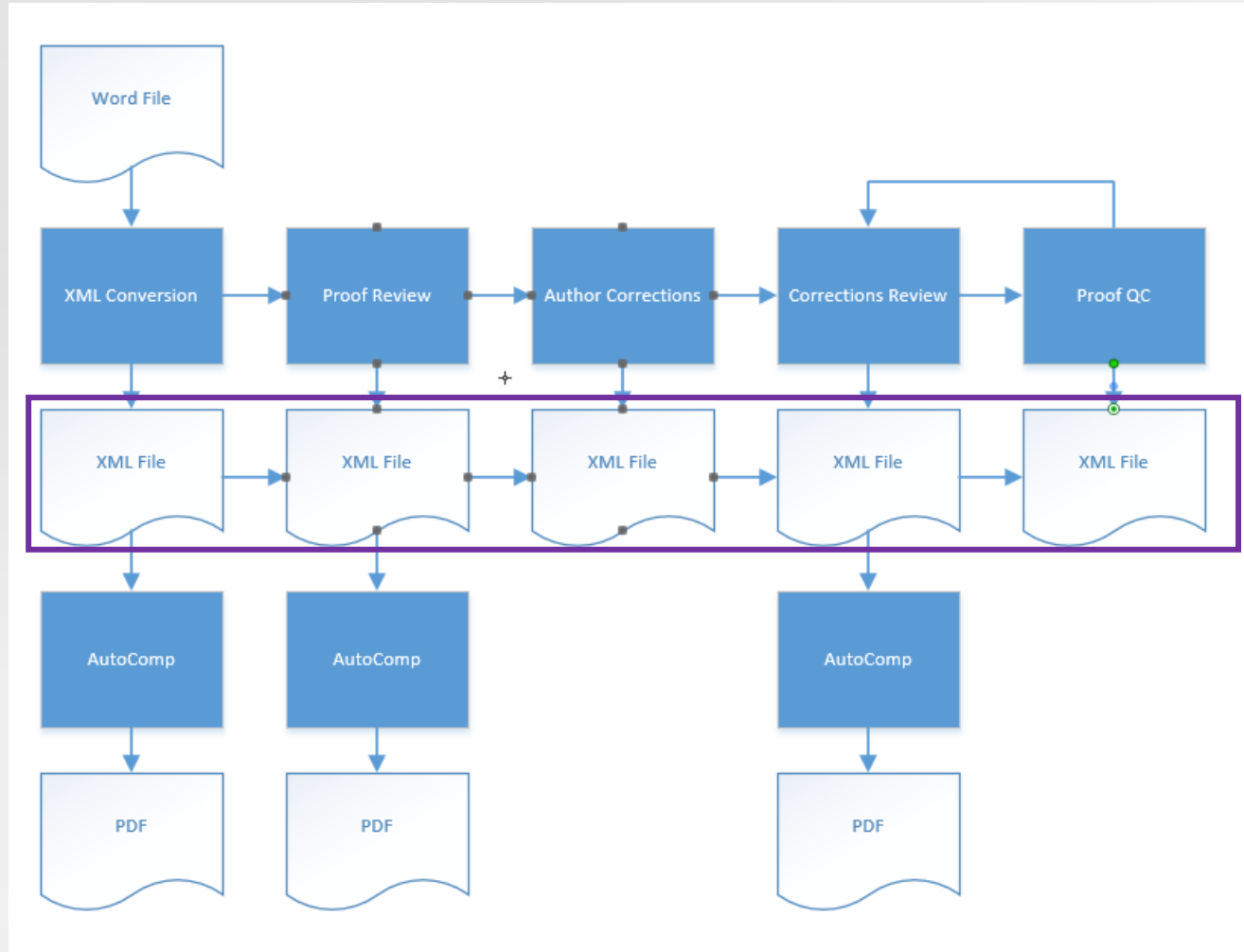
**Conclusions** It is recommended that diagnoses associated with CF be established by evaluation of CFTR function with a variant identified in the Clinical and Functional Translation of CFTR. Newborns with a high immunoreactive sweat chloride and genetic testing may be designated CFTR-related metabolic diagnosis; these terms are now merged and equivalent, and a positive, inconclusive diagnosis may be used. *International Society for*

```
<?xml version="1.0" encoding="UTF-8"?><!DOCTYPE article PUBLIC "-//NLM//DTD JATS (Z39.96) Journal Publishing DTD v1.1 20151215//EN" "JATS-journalpublishing1.dtd"><article article-type="research-article" dtd-version="1.1" xml:lang="EN" xmlns:mml="http://www.w3.org/1998/Math/MathML" xmlns:xlink="http://www.w3.org/1999/xlink"><front><journal-meta><journal-id journal-id-type="publisher-id">sjs</journal-id><journal-title-group><journal-title>sjs</journal-title><abbrev-journal-title abbrev-type="full">sjs</abbrev-journal-title></journal-title-group><issn pub-type="ppub">1111-1111</issn><publisher><publisher-name>sjs</publisher-name></publisher></journal-meta><article-meta><article-id pub-id-type="publisher-id">1</article-id><article-id pub-id-type="doi">10.1175/x</article-id><article-categories><subj-group subj-group-type="heading"><subject>ARTICLES</subject></subj-group></article-categories><title-group><article-title>Discovery of Dynein and its Properties: A Personal Account</article-title></title-group><contrib-group><contrib contrib-type="author"><name><surname>Gibbons</surname><given-names>R.</given-names></name></contrib><aff>Molecular and Cell Biology Department, University of California Berkeley, Berkeley, California</aff></contrib-group><pub-date pub-type="ppub"><month>00</month><year>0000</year></pub-date><volume>00</volume><issue>00</issue><fpage>00</fpage><lpage>00</lpage><history><date date-type="received"><year>0000</year></date><date date-type="received"></date><date date-type="accepted"><year>0000</year></date><date date-type="accepted"></date><date date-type="publishedonline"></date></history><permissions><copyright-statement>Copyright © 2009; SJS</copyright-statement><copyright-year>0000</copyright-year></permissions><self-uri xlink:title="pdf" xlink:href="1.pdf"/><abstract><title>Abstract</title><p>This memoir describes my life as a
```

# Traditional Workflow



# XML Workflow



# LiXuid: The Aries XML Editor

- Content Editing

The screenshot displays the LiXuid XML editor interface. At the top, there is a blue navigation bar with tabs for 'HOME', 'SECTIONS', 'INSERT', 'TITLE PAGE', and 'PRODUCT & BIBLIOGRAPHY'. Below this is a toolbar with various editing tools like undo, redo, bold, italic, underline, and a search bar. The main editing area contains text and a mathematical equation. The text discusses posterior model state  $q_{t+1}$  and preposterior updating. The equation (9) is:

$$\sum_{z_t} \bar{P}(z_t | q_t) V[x_t, q'_t] - V[x_t, q_t],$$

where  $q_t$  is updated to  $q'_t$  based on data  $z_t$  with distribution  $\bar{P}(z_t | q_t) = \sum_k q_t(k) P_k(z_t)$  (see Equation (2)). The first term in Equation (9) is an average optimal valuation from Equation (8) resulting from the updating of the model state with external data. The second term is an optimal valuation based on the current system and model state. The difference expresses the marginal value expected with new sample information. EVSI can be seen to be state-dependent, in that the value given by the comparison in Equation (9) is conditional on the particular combination  $(x_t, q_t)$  of system and model states. That is, different combinations of system and model states can produce different values.

The use of passive adaptive management in EVSI proceeds in much the same way, except that

On the right side, there is a vertical sidebar with tabs for 'OUTLINE', 'COMMENTS', 'QUERIES', and 'IMPROVE'. The 'OUTLINE' tab is active, showing a hierarchical table of contents:

- Front matter
- Article body
  - 1 Decisions, returns, and uncertainty
  - 2 Decision making under structural uncertainty
    - Table 1 Notation used to characterize dynam...
    - 2.1 System dynamics.
    - 2.2 Decision making.
    - 2.3 Propagating uncertainty.
    - 2.4 Valuation.
    - 2.5 Learning-based management.
      - Figure 1
  - 3 Combining internal and external learning in EVSI
    - 3.1 Example: Habitat management for the Flo...
      - Figure 2

At the bottom, there is a breadcrumb trail: 'Article > Front matter > Article meta > Titles > Article title'. In the bottom right corner, there is a notification: '24 improvements' and the Aries systems logo.

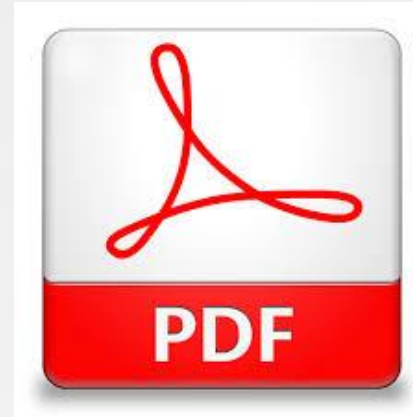
# LiXuid: Auto-Composition

XML

+

=

Graphics



# Aries and XML

 **Editorial Manager<sup>®</sup>**

=

**Metadata**



# EM Meta $\cong$ JATS Meta

- Back to ORCID

```
<contrib-id contrib-id-type="orcid"  
authenticated="true">http://orcid.org/0000-0002-6046-2077</contrib-id>
```

- *caveat*
  - Corresponding Author  $\neq$  Corresponding Author

Aries and XML

 **ProduXion Manager**

=

**Workflow**

Aries and XML

LiXuid  
=  
Content



# Aries and XML

Metadata +  
Workflow +  
Content =

Complete Production Solution



# Future: Peer Review

The screenshot displays a software interface for peer review. At the top, there is a navigation bar with tabs: HOME, SECTIONS, INSERT, TITLE PAGE, PRODUCT & BIBLIOGRAPHY, and TABLE. Below this is a toolbar with various editing tools like bold, italic, underline, and mathematical symbols. The main content area is split into two panes. The left pane contains a table with 8 rows of numerical data. The right pane is titled 'Review' and contains a list of comments from 'Jonathan von Longnamington' dated May 13, 2019. The comments are: 'The data here requires an additional control for wind shear due t...', 'In the Introduction, you failed to cite my ground-breaking study ...', 'Unclear what this equation demonstrates. It is, however, rather e...', and 'The sample size is too small to justify this conclusion.' The last comment is highlighted in yellow and has a reply button.

3	1.0	1	641.76	2	641.06	0.70
4	0.0	3	768.60	1	768.36	0.24
4	0.5	3	702.57	3	704.53	-1.96
4	1.0	2	640.74	2	640.43	0.31
5	0.0	3	766.48	3	766.74	-0.26
5	0.5	3	700.22	3	702.77	-2.55
5	1.0	2	638.71	2	638.85	-0.13

More data from annual monitoring should produce increased value on average over the long term, a result borne out from long-term simulations that account for the likelihood of occurrence for different states (Supporting Information). Nonetheless, the advantage of annual monitoring over biennial monitoring appears to be very small in this example, probably because of the strong relationship between states in successive years. This confirms the intuitive result that there is little to be gained from the frequent monitoring of slowly changing ecosystems.

**Review**

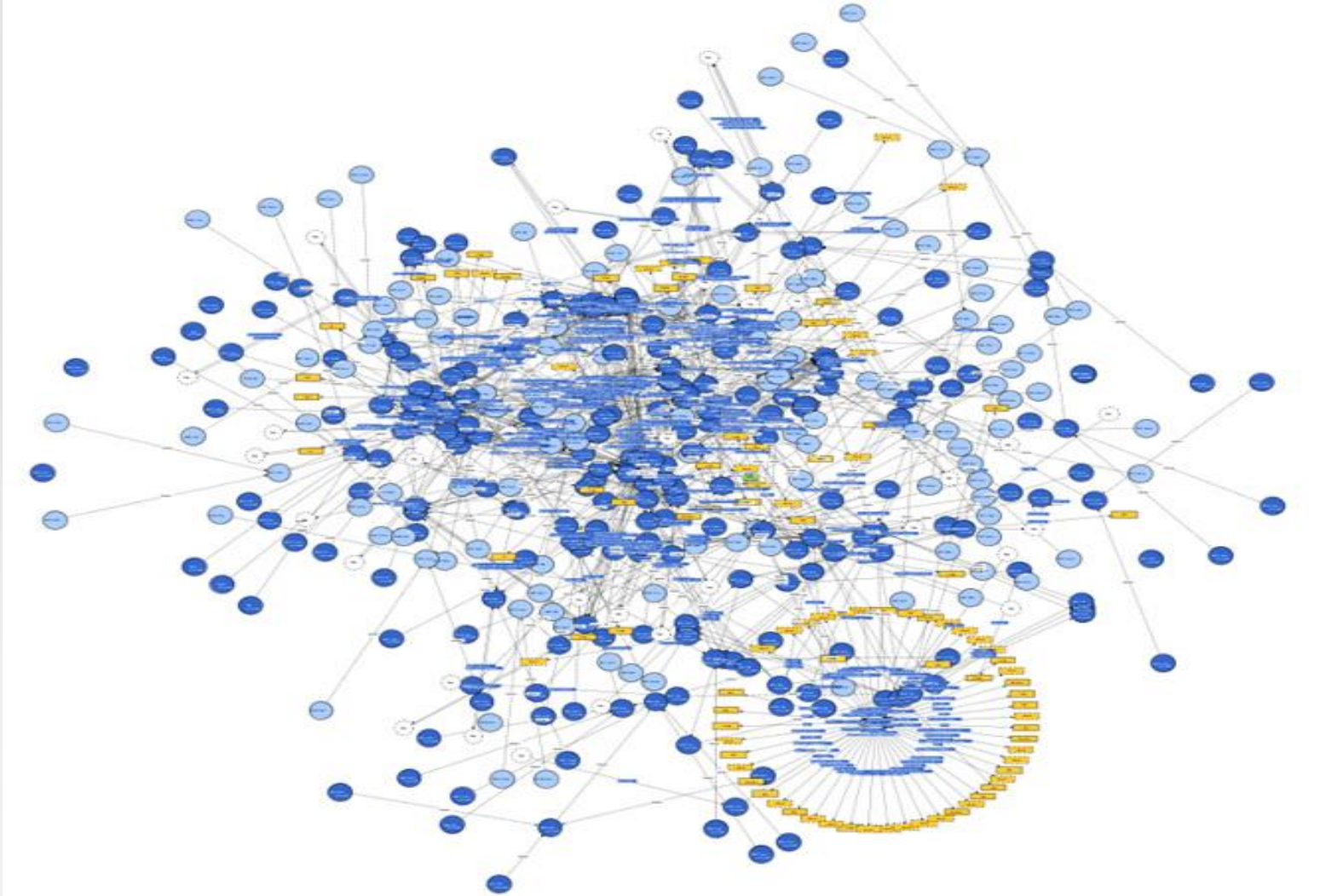
Jonathan von Longnamington - May 13, 2019 7:22:31 PM  
**Commented**  
The data here requires an additional control for wind shear due t...

Jonathan von Longnamington - May 13, 2019 7:23:53 PM  
**Commented**  
In the Introduction, you failed to cite my ground-breaking study ...

Jonathan von Longnamington - May 13, 2019 7:26:49 PM  
**Commented**  
Unclear what this equation demonstrates. It is, however, rather e...

Jonathan von Longnamington - May 13, 2019 7:25:34 PM  
**Commented**  
The sample size is too small to justify this conclusion.

# Future: Semantic Enrichment



# Questions?