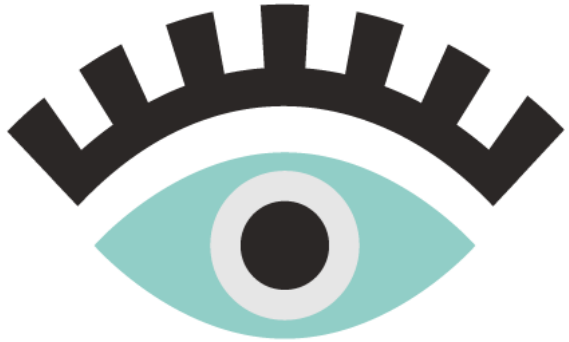
EEMUG 2020

#EEMUG2020

EEMUG 2020

XML Bootcamp

Let´s Mark up Things ☺



#EEMUG2020

Charles O´Connor
Business Systems Analyst
coconnor@ariessys.com

Tayyip Sahin
Account Coordinator
tsahin@ariessys.com

Agenda

- What is XML?
- XML is not!
- Binary vs. Text
- What is Markup?
- XML-Syntax
- XML Contains – Elements and Attributes
- XML Links
- XML vs HTML
- Linking in XML and HTML
- XML-Content → Presentation
- XML is a Tree
- The Pieces of an XML
- An XML Document is Defined by a DTD
- The JATS DTD
- Applications and Functions of JATS in EM
- Conclusion

What is XML?

- XML stands for eXtensible Markup Language
- XML is a markup language much like HTML
- XML was designed to describe data
- XML tags are not predefined in XML. You must define your own tags
- XML uses a Document Type Definition (DTD) to describe the data
- XML with a DTD or XML Schema is designed to be self-descriptive

XML is not!

- A proprietary binary format like...
- WORD or PDF

or

- *A replacement for HTML*, but HTML can be generated from XML.
- *A presentation format*, but XML can be converted into one.
- *A programming language*, but it can used with almost any language
- *A network transfer protocol*, but XML may be transferred over a network
- *A database*, but XML may be stored into a database

Binary vs. Text

- Binary formats are platform-dependent, have firewalls, they are hard to debug, and inspecting the file can be a difficult task.
- XML is text-based and is not bound to any of the above requirements.

- XML is a series of tags that represent some form of data. Here is a very simplistic XML file:

```
<article><data message="Well, hello there!"/></article>
```

Binary is a series of ones and zeroes. Here is the exact same XML file in binary:

00111100 01110010 01101111 01101111 01110100 00111110 00111100 01100100
01100001 01110100 01100001 00100000 01101101 01100101 01110011 01110011
01100001 01100111 01100101 00111101 00100010 01010111 01100101 01101100
01101100 00101100 00100000 01101000 01100101 01101100 01101100 01101111
00100000 01110100 01101000 01100101 01110010 01100101 00100001 00100010
00101111 00111110 00111100 00101111 01110010 01101111 01101111 01110100
00111110 00001010

What is Markup?

- Information added to a document to enhance its meaning in certain ways
- Set of symbols that can be placed in the text document to demarcate and label the parts or it.

- **Like HTML**

<h1>This is a first-level section heading</h1>
<h2>This is a second level section heading</h2>

<p>This is a paragraph of the text<p>

- **Or Markdown**

This is a first-level section heading
This is a second-level section heading

This is also a paragraph of a text just marked down

XML Syntax

- XML declaration is the first statement
- All XML elements must have a **closing tag**
- XML tags are **case sensitive**
- All XML elements must be **properly nested**
- All XML documents must have a **root tag**
- **Attribute** values must always be **quoted**
- With XML, white space is preserved
- Comments in XML: <!-- This is a comment -->
- Certain characters are reserved for **parsing**

XML Contains

- **Elements**

What exactly are elements?

```
Please see our companion paper,
<associated-article>Sahin et al, 2019</associated-article>.
```

- **Attributes**

What exactly are attributes?

```
<associated-article id="ra1" associated-article="companion"
ext-link-type="doi" xlink:href="10.1515/tsdemo.54321">
Sahin et al, 2019)</associated-article>
```

XML Links

Sample link in XML

```
<xref ref-type="fig" rid="figure1">Fig. 1a - 1c</xref>
```

Reference Target

```
<fig id="figure1"><label>Figure 1</label>
<caption><p>This figure represents the Milkyway.</p></caption>
<graphic xlink:href="demots.00000001_figure1"/></fig>
```

XML vs HTML

- **HTML Describes How Text Should be Displayed**

```
<h1>The Daltons</h1>
        <ul>
                <li>Joe Dalton</li>
                <li>Averell Dalton</li>
        </ul>
```

- **XML Describes the Meaning**

```
<article-title> The Daltons </article-title>
        <contrib>
                <given-names>Joe</given-names>
                <surname>Dalton</surname>
        </contrib>
        <contrib>
                <given-names>Averell</given-names>
                <surname>Dalton</surname>
        </contrib>
```

Linking in XML and HTML

- **Basic Database Link in HTML**

```
<a href="https://www.ncbi.nlm.nih.gov/nlmcatalog/9890633">
```

- **Basic Database Link in XML**

```
<ext-link ext-link-type="gen" xlink:href="9890633">
```

XML-Content → Presentation

- **XML-Content**

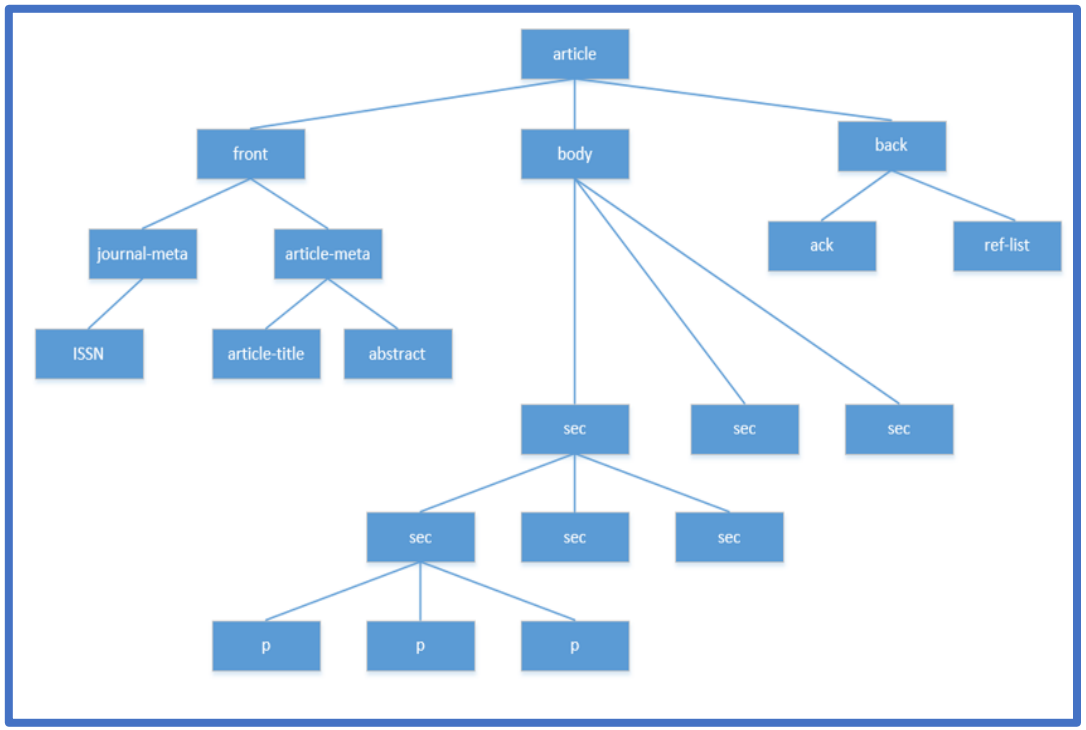
```
<history>
    <date date-type="received">
        <day>22</day><month>12</month><year>2016</year>
    </date>
    <date date-type="accept">
        <day>21</day><month>04</month><year>2017</year>
    </date>
</history>

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

- **Presentation Online**

Received: December 22, 2016; **Accepted:** April 21, 2017; **Published:** May 22, 2017

XML is a Tree

- An article XML document modeled as tree




```xml
<?xml version="1.0" encoding="UTF-8"?>
<article>
  <front category="front">
    <journal-meta>
        <ISSN>xxxxxxx</ISSN>
    </journal-meta>
    <article-meta>
        <article-title>Milky Way</article-title>
        <abstract>Summary of the Milkly Way</abstract>
    </article-meta>
  </front>

  <body category="body">
    <sec>
        <sec>
            <p>Moon</p>
            <p>Earth</p>
            <p>Sun</p>
        </sec>
        <sec>Alpha Centauri</sec>
        <sec>Andromeda</sec>
    </sec>

    <sec>Milky Way</sec>
    <sec>Solar System</sec>
  </body>

  <back category="back">
    <ack>xxxxxxxx</ack>
    <ref-list>xxxxxxxxx</ref-list>
  </back>
</article>
```

The Pieces of an XML

- There are 3 components for XML content
    1. The XML document
    2. DTD (Document Type Declaration)
    3. XSL (Extensible Stylesheet Language)

An XML Document is Defined by a DTD

- DTD is short for Document Type Definition.
- The DTD establishes the vocabulary for one XML application.
- What elements and attributes can appear in a document?
- What is the order of the defined elements.
- What can appear in elements
    - Only other elements?
    - Only text?
    - Text and other elements?
- DTDs include JATS, NLM, BITS, DocBook, DITA, TEI, etc

The JATS DTD

- Standard developed by the U.S. National Library of Medicine
- The first version was released in March 2003
- In July 2012 the Journal Article Tag Set became a NISO (National Information Standards Organization) standard.
- JATS is the standard for journal articles in scholarly publishing – not only science, technology and medicine but also other branches.

Applications and Functions of JATS 1

- **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.

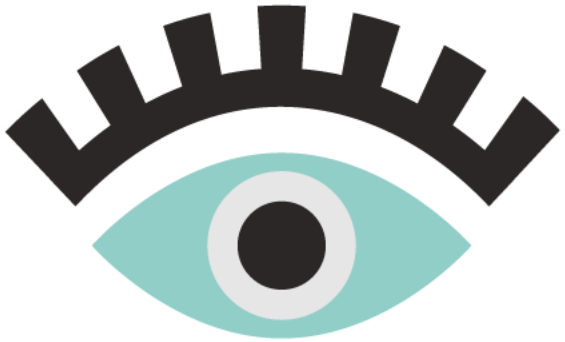
Applications and Functions of JATS 2

- **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.

Conclusion

- XML is a **self-descriptive language**
- XML is a powerful language to describe **structure data** for **web application**
- XML is currently applied in many fields not just in scholarly publishing
- Many vendors already supports or will support XML
- XML Documents can be **validated** through the use of **DTD** documents
- XML impacts **B2B data exchanges**, legacy system integration, web page development, database system integration.

Questions?



#EEMUG2020

XML Bootcamp

The Impact of JATS/XML on Scholarly Publishing

Charles O´Connor
Business Systems Analyst
coconnor@ariessys.com

Tayyip Sahin
Account Coordinator
tsahin@ariessys.com

A Bit of History . . . SGML

- SGML: Standard Generalized Markup Language
- Includes familiar angle brackets, <tagged>but</tagged> the syntax is more complex
  - Tags can be omitted (if unambiguous)
  - Null End Tags: “<italic/cheese/” = <italic>cheese</italic>
  - Documents may contain other documents
  - Etc.
- XML is a *subset* of SGML (as was HTML, until HTML5)

The Rise of JATS

- Online-only journals and the need for archiving
- PDFs? Nooooooooooooooooooo!
  - Binary formats go out of style: Betamax v. VHS
  - Less accessible metadata
  - Less machine readable
- Who remembers ISO 12083:1994, Electronic Manuscript Preparation and Markup?
- Proprietary XML DTDs

XML-Related Technologies

- XPath: Query language for finding stuff in an XML document
  - EX: article/body/sec[1]/sec[1]/p[3]
- XSLT: Transforms XML into HTML, Text, other XML, etc.
- XQuery: Like SQL, but for XML. Transforms information in XML into other data formats
- Schematron: Rule-based validation language

```
<sch:rule context="pub-date" role=“warning">
    <sch:report test="year &gt; 2020">The year is in the future.</sch:report>
</sch:rule>
```

JATS/XML: What Is It Good For?

- Metadata Initiatives
- Semantic Tagging
- Production Workflows

Metadata Initiatives: ORCID

- Unique identifier for contributors
- Disambiguates “Jane Smith” and “Jane Smith”
- JATS example:

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

Metadata Initiatives: CHORUS

```
<funding-group specific-use="Crossref Funding Data">
  <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>
```

(Example from JATS 1.1)

Metadata Initiatives: CHORUS

```
<license>
    <ali:license_ref xmlns:ali="https://www.niso.org/schemas/ali/1.0/"
      specific-use="am" start_date="2020-01-23">
      https://creativecommons.org/licenses/by/4.0/</ali:license_ref>
</license>
```

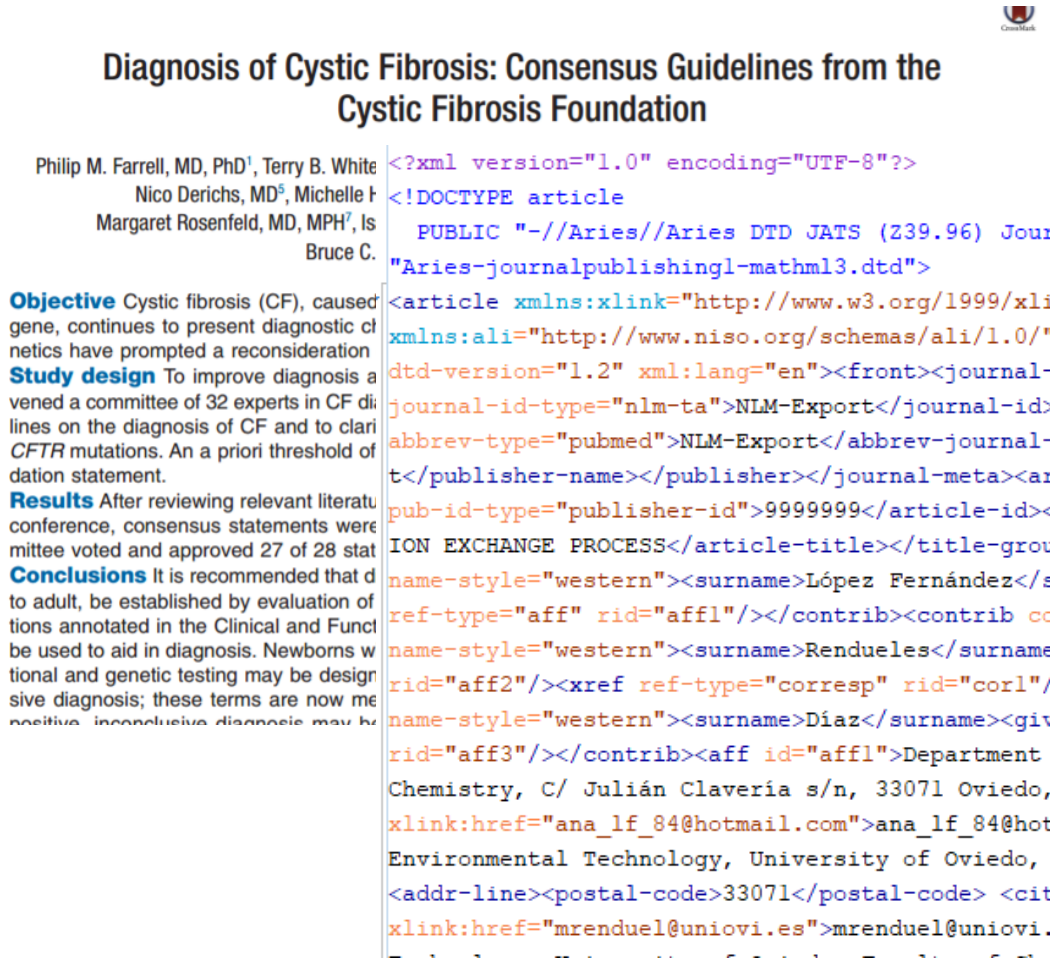
Semantic Tagging: Vocab Attributes

```
<contrib>
    <string-name>
        <given-names>Dan</given-names>
        <surname>Green</surname>
    </string-name>
    <role vocab="credit" vocab-identifier=
"http://dictionary.casrai.org/Contributor_Roles"
        vocab-term="Conceptualization"
        vocab-term-identifier=
"http://dictionary.casrai.org/Contributor_Roles/Conceptualization">
        Conceptualization</role>
</contrib>
```

Semantic Tagging: Vocab Attributes

```
<article-version vocab="JAV"
  vocab-identifier="http://www.niso.org/publications/rp/RP-8-2008.pdf"
  article-version-type="VoR"
  vocab-term="Version of Record">Published version
</article-version>
```

Production Workflows



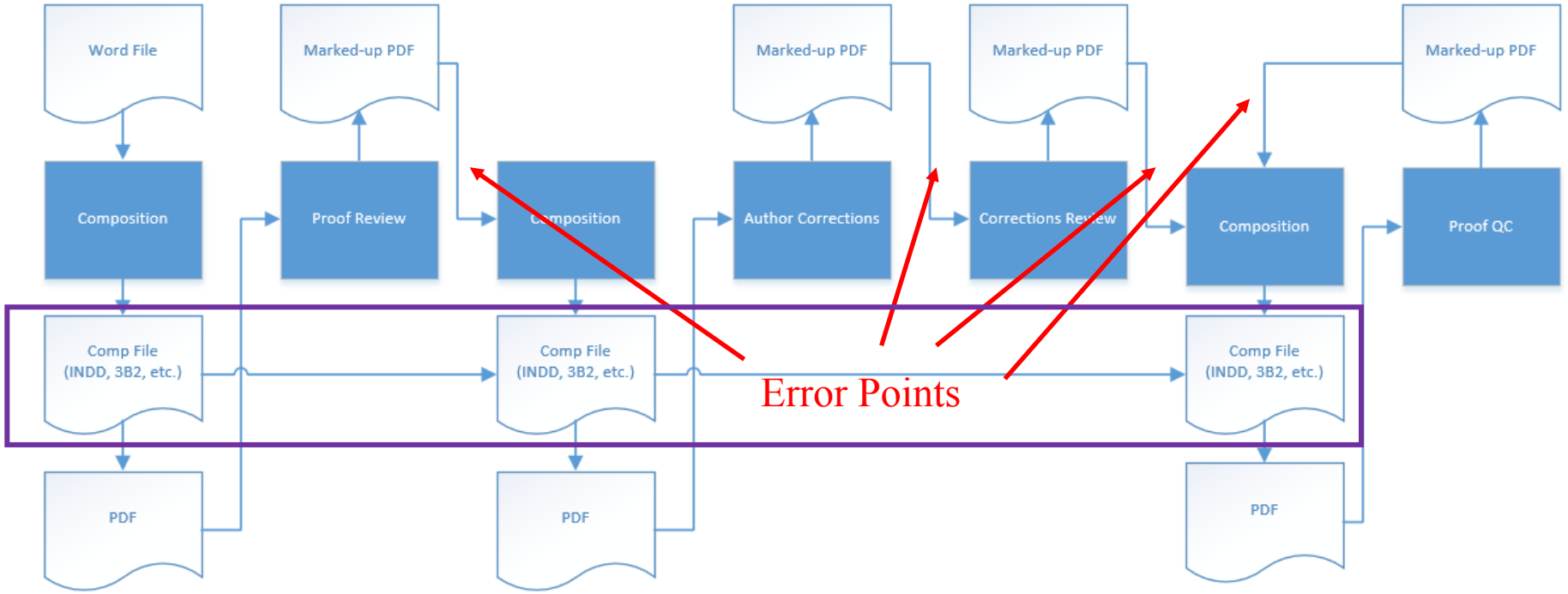
Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation

Philip M. Farrell, MD, PhD[1], Terry B. White
Nico Derichs, MD[5], Michelle H
Margaret Rosenfeld, MD, MPH[7], Is
Bruce C.

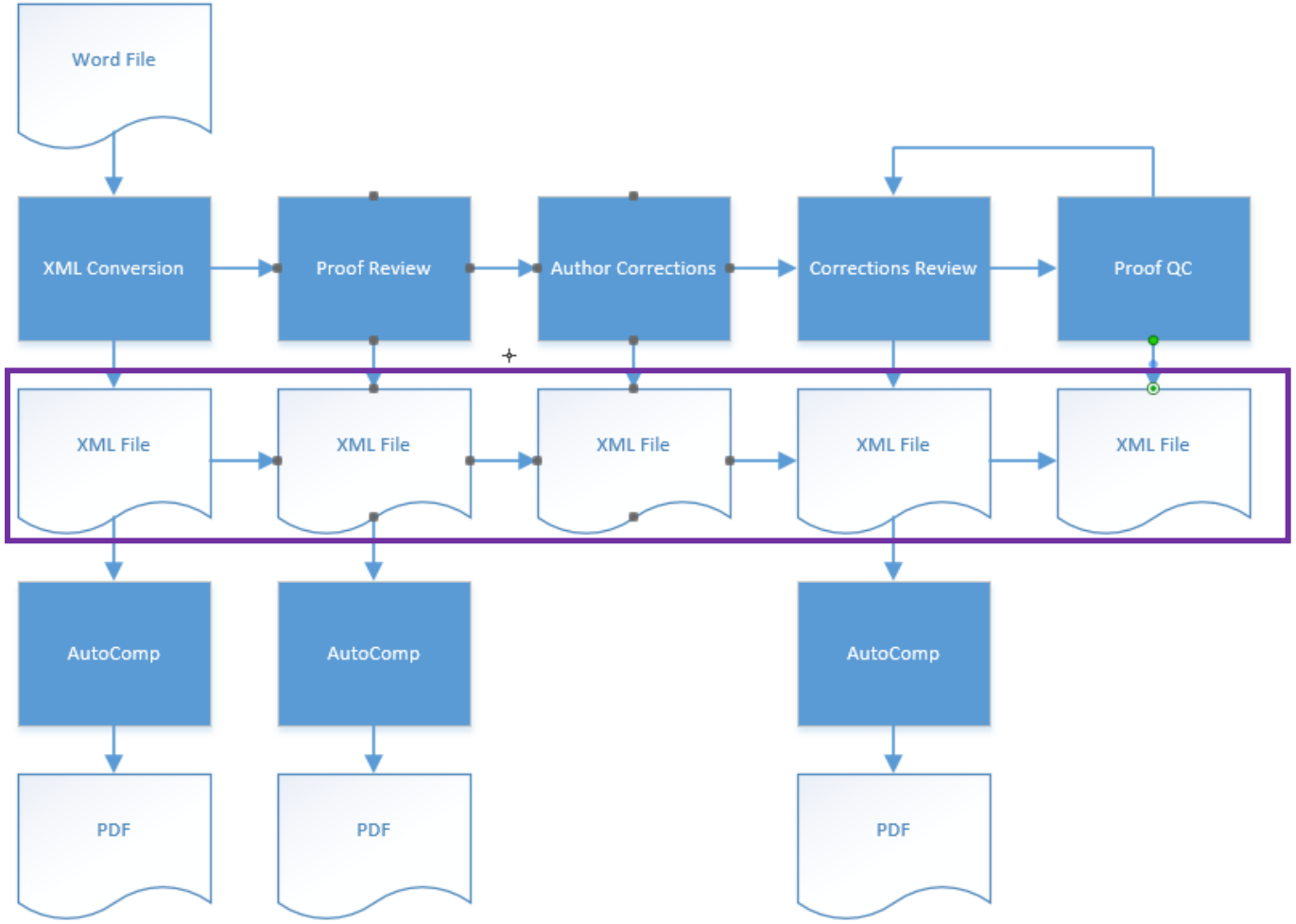
Objective Cystic fibrosis (CF), caused
gene, continues to present diagnostic ch
netics have prompted a reconsideration
Study design To improve diagnosis a
vened a committee of 32 experts in CF di
lines on the diagnosis of CF and to clari
CFTR mutations. An a priori threshold of
dation statement.
Results After reviewing relevant literatu
conference, consensus statements were
mittee voted and approved 27 of 28 stat
Conclusions It is recommended that d
to adult, be established by evaluation of
tions annotated in the Clinical and Funct
be used to aid in diagnosis. Newborns w
tional and genetic testing may be design
sive diagnosis; these terms are now me
positive, inconclusive diagnosis may be

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE article
  PUBLIC "-//Aries//Aries DTD JATS (Z39.96) Journal Publishing DTD with MathML3 v1.2 20190208//EN"
"Aries-journalpublishing1-mathml3.dtd">
<article xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:mml="http://www.w3.org/1998/Math/MathML"
xmlns:ali="http://www.niso.org/schemas/ali/1.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" article-type="research-article"
dtd-version="1.2" xml:lang="en"><front><journal-meta><journal-id journal-id-type="publisher-id">NLM-Export</journal-id><journal-id
journal-id-type="nlm-ta">NLM-Export</journal-id><journal-title-group><journal-title>NLM-Export</journal-title><abbrev-journal-title
abbrev-type="pubmed">NLM-Export</abbrev-journal-title></journal-title-group><issn>0317-8471</issn><publisher><publisher-name>NLM-Expor
t</publisher-name></publisher></journal-meta><article-meta><article-id
pub-id-type="publisher-id">9999999</article-id><title-group><article-title>COMPETITION OF SALTS WITH SULFAMETHOXAZOLE IN AN ANIONIC
ION EXCHANGE PROCESS</article-title></title-group><contrib-group><contrib contrib-type="author"><name
name-style="western"><surname>López Fernández</surname><given-names>Ana María</given-names></name><degrees>PhD</degrees><xref
ref-type="aff" rid="aff1"/></contrib><contrib contrib-type="author" corresp="yes"><name
name-style="western"><surname>Rendueles</surname><given-names>Manuel</given-names></name><role>Lecturer</role><xref ref-type="aff"
rid="aff2"/><xref ref-type="corresp" rid="cor1"/></contrib><contrib contrib-type="author"><name
name-style="western"><surname>Díaz</surname><given-names>Mario</given-names></name><role>Prof.</role><xref ref-type="aff"
rid="aff3"/></contrib><aff id="aff1">Department of Chemical Engineering and Environmental Technology, University of Oviedo, Faculty of
Chemistry, C/ Julián Clavería s/n, 33071 Oviedo, Spain. E.mail: <email
xlink:href="ana_lf_84@hotmail.com">ana_lf_84@hotmail.com</email></aff><aff id="aff2">Department of Chemical Engineering and
Environmental Technology, University of Oviedo, Faculty of Chemistry, C/ Julián Clavería s/n,
<addr-line><postal-code>33071</postal-code> <city>Oviedo</city></addr-line>, <country>Spain</country>. E-mail <email
xlink:href="mrenduel@uniovi.es">mrenduel@uniovi.es</email></aff><aff id="aff3">Department of Chemical Engineering and Environmental
```

Traditional Workflow

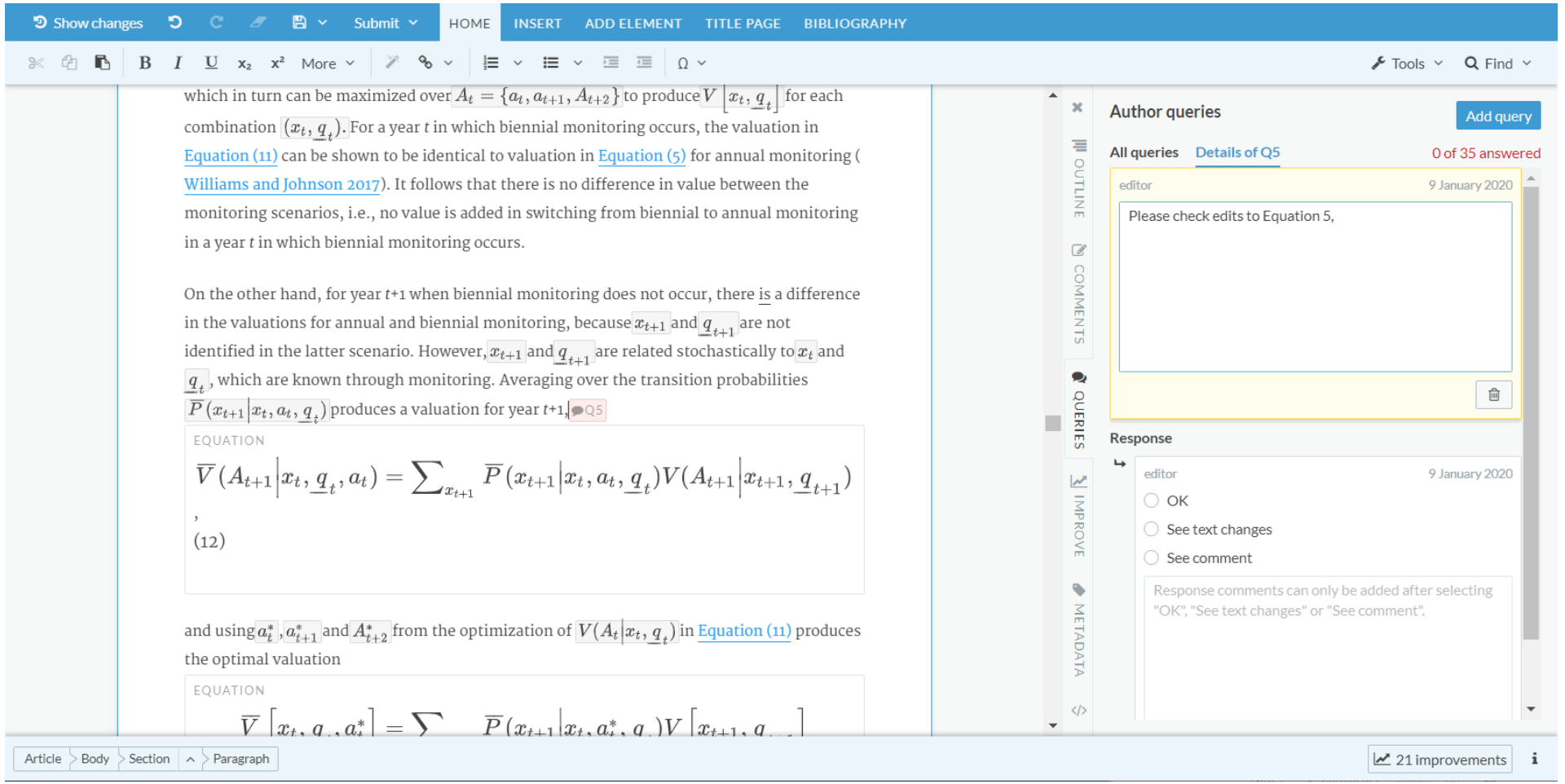


XML Workflow



LiXuid: The Aries XML Editor

- Content Editing



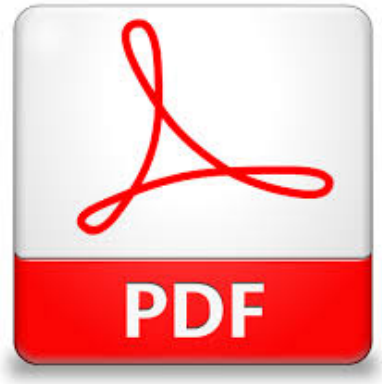
LiXuid: Auto-Composition

XML

+

Graphics

=



Aries and XML

=

Metadata

EM Meta ≅ 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 ≠ Corresponding author

Aries and XML

=

Workflow

Aries and XML

LiXuid Manuscript™

=

Content

Aries and XML

Metadata +

Content +

Workflow =

Complete Workflow Solution

Questions?