

XML 101 for Journal Production Editors

MODERATOR:

Julie L Nash

Senior Partner
J&J Editorial, LLC
Cary, North Carolina

SPEAKERS:

Michael Casp

Director of Business
Development, Production
Services Coordinator
J&J Editorial, LLC
Cary, North Carolina

Karie Kirkpatrick

Associate Publisher, Digital
American Physiological Society
Rockville, Maryland

REPORTER:

Heather DiAngelis

Associate Publications Director,
Transportation Research Board
National Academies of Sciences,
Engineering, and Medicine
Washington, DC

This session provided contextual overview, historical background, and best practice suggestions for XML novices in publishing.

Michael Casp, Director of Business Development and Production Services Coordinator at J&J Editorial, began the session with a definition of extensible markup language (XML) and gave basic visual examples of how it is both a human-readable and machine-readable language. This mark-up language is everywhere—even in Microsoft Word. His boiled-down definition was “XML is labels.” He went on to explain various versions of XML document type definitions (DTDs) and various ways to organize them, such as NLM, JATS, and custom DTDs for content management systems. Journal article tag suite (JATS) is the most commonly used DTD for scholarly journal content;

it has become the de facto XML standard in the scholarly publishing industry.

Casp also discussed how peer review systems such as Editorial Manager collect metadata from authors and editorial staff and then convert them to XML files for transmittal. Afterward, a manuscript document is tagged either before or after copyediting, mostly via automation with human review, to create a new XML file of the text, tables, and metadata; as he noted, almost every item relevant to the article except the figures will be contained within the XML. He continued by discussing what one can do with this XML, such as generating PDFs, creating webpages, or indexing (e.g., through Crossref). Casp ended by remarking on the importance of quality checking XML with automated tools (for validity) and a good eye (for accuracy) to find errors that can cause ingestion failures or downstream problems. His final takeaway was that XML is flexible, findable, and fast.

Karie Kirkpatrick, Associate Digital Publisher for the American Physiological Society, presented second. She discussed the benefits of JATS, including its universality, conversion tools, and large and helpful community of users and resources. She discussed different types of journal and article metadata commonly used in JATS, and then she demonstrated how articles can be tagged with eXtensible Markup Language (XML) and converted to XML after validation against the JATS DTD. From there, incorrect elements such as typos in the data or incorrect ISSNs can cause bad XML even when it is valid.

Kirkpatrick then defined Schematron, a rules-based validation language that creates rules and output messages and can catch recurring mistakes that a DTD can't detect. A helpful community of creators exists to help other production editors create and discuss these Schematrons, which can be run in-house or at a compositor and can vary in size from one rule to thousands. Kirkpatrick ended her presentation with a reminder that Schematrons have great downstream implications, as they help avoid XML redeposits and ensure indexers and archivists don't receive bad XML.

A question-and-answer period followed the presentations, with audience questions about XML-viewing programs and cleaning up XML errors before and after publication.

WHAT IS XML?

extensible Markup Language

XML Definition

- a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.
- a metalanguage which allows users to define their own customized markup languages, especially in order to display documents on the Internet.

```
<?xml
<title>XML test</title>
<text type="test">
  <body>
    <p>
      Though this is a very pared
      <br />
      down XML document, it nonetheless
      <br />
      provides an example of how an XML
      <br />
      document displays on the web without
      <br />
      the intercession of a stylesheet or
      <br />
      other conversion program.
    </p>
  </body>
</text>
</xml>
```