|  |
| --- |
| NATIONAL LIBRARY OF MEDICINE |
| A Journal Article Tag Suite (JATS) extension to support NISO’s Recommended Practices for Online Supplemental Journal Article Materials |
|  |
| Spring Project Report |
|  |
| August 5, 2013  Karen Gutzman  Project Leaders:  Jeff Beck and Kimberly A. Tryka |
|  |

|  |
| --- |
|  |

Contents

[Abstract 3](#_Toc363475994)

[Background: 3](#_Toc363475995)

[Objective: 3](#_Toc363475996)

[Methods: 3](#_Toc363475997)

[Outcomes: 3](#_Toc363475998)

[Introduction 4](#_Toc363475999)

[Markup 4](#_Toc363476000)

[XML 4](#_Toc363476001)

[DTD 5](#_Toc363476002)

[Uses of XML 6](#_Toc363476003)

[Document Modeling 7](#_Toc363476004)

[JATS Standard 7](#_Toc363476005)

[Supplementary Material 8](#_Toc363476006)

[NISO Recommendation 9](#_Toc363476007)

[Preliminary Work 10](#_Toc363476008)

[JATS Supplementary Material Structure 10](#_Toc363476009)

[NISO RP Supplementary Material Structure 11](#_Toc363476010)

[13](#_Toc363476011)

[An Extension 13](#_Toc363476012)

[The JATS Extension (EXT) 13](#_Toc363476013)

[Major Revisions 14](#_Toc363476014)

[Revisions to the NISO RP 14](#_Toc363476015)

[Revisions to the JATS Journal Publishing Model 15](#_Toc363476016)

[Tag Libraries 15](#_Toc363476017)

[EXT Supplementary Material Structure 16](#_Toc363476018)

[Recommendations 17](#_Toc363476019)

[Acknowledgements 17](#_Toc363476020)

[References 17](#_Toc363476021)

[Appendix 1: Glossary of Terms 19](#_Toc363476022)

[Appendix 2. Example of XML (Recipe) 22](#_Toc363476023)

[Appendix 3. Example of DTD (Recipe) 24](#_Toc363476024)

[Appendix 4: Gantt chart of Project Duration 25](#_Toc363476025)

[Appendix 5. Schematic representation of XML Structures 26](#_Toc363476026)

[29](#_Toc363476027)

[Appendix 6. Minor Revisions between NISO RP and JATS 30](#_Toc363476028)

[Appendix 7. Listing of Examples showing JATS EXT markup 31](#_Toc363476029)

# Abstract

## Background:

As online publication became common, publishers capitalized on the ability to provide users with more information than what was originally supplied by print publications. The additional materials included tables, graphs, movies and other materials that were supplemental to the original article. Publishers developed different policies for making the material available to users, with different formats, displays, and preservation techniques employed.

In January 2010, the National Information Standards Organization (NISO) and the National Federation of Advanced Information Services (NFAIS) co-sponsored a roundtable discussion that eventually led to the creation of the NISO *Recommended Practices for Online Supplemental Journal Article Materials* (referred to as: NISO RP) in January 2013*.*

## Objective:

The purpose of the project was to create an extension of the Journal Article Tag Suite (JATS) which is a common format used by the National Center for Biotechnology Information (NCBI) to describe the intellectual content of journal articles in order to support the NISO RP.

## Methods:

The associate fellow learned basic XML, the theory and practice of document modeling, the Document Type Definition (DTD) schema language, and the practical use of the XML editor, oXygen. The fellow became familiar with JATS and the NISO RP through basic tagging exercises. The team had discussions on the current JATS model and issues that led to the creation of the recommendation. They reviewed sample articles and studied the structure of the non-normative DTD from the recommendation. They then built an extension of the JATS supplementary material model that supports the NISO RP.

## Outcomes:

The team created an extension to JATS that supports the recommendation. This extension will be made available with other JATS extensions on the JATS non-normative documentation site at <[http://jats.nlm.nih.gov/](http://jats.nlm.nih.gov/index.html)>. A preliminary tag library for this JATS extension can be found at <<http://www.ncbi.nlm.nih.gov/staff/beck/sme/index.html>>.

# Background

The main outcome of this project was an extension to the Journal Article Tag Suite (JATS) that better supports supplementary material by using the National Information Standard’s Organization’s (NISO) Recommended Practices for Journal Article Materials (referred to as the NISO RP). There were several areas that needed to be addressed in order for me to participate in the work being done, and for the extension to attain its full potential. I explored how markup languages, such as Extensible Markup Language (XML), are best utilized to support the needs of the user. I learned how XML models can be created for specific types of documents through a process called document modeling, and how to create a basic set of rules which require all users to follow the prescribed model. In addition, I gained an understanding of the purpose of JATS and the issues surrounding supplementary material. I will address each of these areas in order to best represent the boundaries of the project.

## Markup

Whenever an author writes anything, he or she “marks it up.” Through the use of punctuation and spaces between letters, markup allows users to clarify the expression of words. Editors used special markup to represent where authors needed to add new paragraphs, bold text or make other edits. New types of markup and processing were created with the advent of electronic text-processing systems in the 1960s and 70s. When a journal article is displayed, either in print or electronically, it still contains the markup (punctuation and spaces between letters) that we as humans need in order to process and understand the content of the text. However, in an electronic format, the article also contains additional types of markup which allow the computer to process the text (Coombs et al, 1987). Differing systems of electronic markup have evolved into markup languages.

Electronic markup languages can be applied to a document with many outcomes. Markup languages focused on description indicate which areas of text belong to a certain class. An example is markup surrounding individual paragraphs, which when processed allows the computer to separate each paragraph appropriately instead of displaying one big paragraph of text. Markup languages focused on presentation might surround the title of a chapter in a book, and when processed the computer would display the title in a pre-specified format, such as 14 point, Times New Roman font.

## XML

The markup used for this project is Extensible Markup Language (XML). XML is a descriptive markup which focuses on describing the structure and meaning of a document, and does not focus much on the presentation of the document. XML is “extensible” because user determines the types of elements and attributes necessary to describe the document at will (TEI Consortium, 2013). This is unlike other markup languages, such as Hypertext Markup Language (HTML) which provides users with a defined set of elements and attributes which have to be followed in order to have a well-formed, validated HTML document.

Figure 1 provides a simple example of XML markup which has been applied to a message. Elements represent an information object or text in a document and are surrounded by angled brackets. The opening and closing instances of an element are often referred to as a “tags.” Elements are case sensitive, must start with a letter (i.e. a-z or A-Z) and can contain some special characters such as dashes or underscores. Each document has one root element which encloses all other elements. Attributes can be thought of as modifiers of elements; they are found in the opening tag, and often provide information that is not explicitly part of the text. Attribute values can be free text or values, such as terms from a controlled vocabulary which can be enforced using a schema. A glossary of all major terms used in this paper can be found in Appendix 1.

<message lang=“English”>

<title>Directions to the National Library of Medicine</title>

<p>In order to visit you will need to enter the NIH campus through the Visitor Gateway Center. </p>

<p>The security guard will provide you with a map and limited instructions on finding the library. </p>

<p>Exit through the second floor of the Visitor Gateway Center and follow the sidewalk. Cross the street and continue to NLM’s front entrance. </p>

</message>

The element <p> indicates the opening of a paragraph element, and </p> indicates the closing of a paragraph element.

The element <message> is the root element.

The attribute lang has a value of English, indicating the language of the message is English.

**Figure 1.** Example of electronic markup on a message.

However, making up tags “at will” could mean that two people may tag the same document with different elements and attributes. In Figure 1, the <message> element tagged by one person may be considered a <memo> element by someone else. In order to create boundaries or rules when tagging similar document types, XML employs a structure called an XML schema language. An XML schema language is used to express a set of rules to which an XML document must conform to be considered valid according to that schema. There are three main XML schema languages: XML Schema (also called XSD), Document Type Definition (DTD), and Regular Language for XML Next Generation (RELAX NG) (XML Schema Languages, 2013).

## DTD

Document Type Definition (DTD) is one of the three main XML schema languages, and the XML schema language used in this project. A DTD provides a model for the type of document being tagged, including the elements or attributes allowed, the hierarchy of those elements and attributes, as well as their allowed content. One advantage of this type of model is that different people can tag documents using the same DTD and be able to exchange their files with some effort.

In XML, elements have hierarchical relationships to one another through the concept of ancestry. The DTD declares the root element which forms the top level and surrounds the whole document. Sibling elements are found at the same level, and child elements are only found below their parent element. In Figure 1, the root element is <message>, which is also the parent element to child elements <title> and <p>.

If the content inside an element includes text or data, this is noted in the DTD by the term #PCDATA. The DTD also includes instructions on which child elements are allowed inside a parent element, how often a child may be repeated (i.e. number qualifiers) and in what sequence (i.e. sequence qualifiers) the child may appear under a parent element through a system of symbols, as defined in Figure 2.

| **Figure 2. Allowable content inside elements is declared in the DTD through the use of element and sequence qualifiers in the form of symbols.** | |
| --- | --- |
| **Element Number Qualifiers** | |
| Symbol | Definition |
| ? | The element is optional and cannot repeat (0 or 1) |
| + | The element is required and may repeat as many times as necessary (1 or more) |
| \* | The element is optional and may repeat as many times as necessary (0 or more) |
| **Sequence Qualifiers** | |
| Symbol | Definition |
| , | The elements must occur in the order provided (with the elements in the list being separated by a comma) |
| | | The user may choose between elements separated by the pipe |
| & | All elements are available in any sequence the user desires. |

A DTD may be declared internally in an XML document, or declared as an external reference. An example of an externally reference DTD is found in the XML document in Appendix 2. The DTD for the XML document in Appendix 2 is found in Appendix 3.

## Uses of XML

Until the advent of the internet, print documents were the main way of widely communicating research results. Information was gathered by scouring through print documents and then organized and analyzed by hand. With the advent of computers, documents began to be shared through computer systems. Computers can search for, organize and analyze the information contained in documents much more quickly than ever before, and the use of XML can facilitate these tasks. According to Allen Renear’s work on Text Encoding (Renear, 2004), using an XML structure provides several advantages in the areas of formatting, processing, information retrieval, and analysis.

* **Formatting.** XML generally contains very few presentational elements. In order to change the formatting of paragraphs in a document, the rule associated with formatting commands in the descriptive markup can be easily modified.
* **Processing.** XML allows for automated processing of all elements in a document. An example is the use of XML to create a table of contents by an automated pulling and gathering of all title elements from a document. Another example is the creation of an index based on the keyword elements in a document.
* **Information Retrieval.** XML supports information retrieval by allowing the elements inside a document to be treated as fields in a database. An example of this is a user being able to request title elements which contain a particular disease or chemical name, rather than relying on a full text search of the entire document.
* **Analysis.** XML supports analytical procedures by “allowing the computer to distinguish text, data, directions and so on, so that it can be pulled apart from the whole and analyzed (Renear, 2004).”

## Document Modeling

Document modeling is the process of examining existing structures to identify structures and features in the text, and then to classify, and note the relationships between those pieces. In layman’s terms, document modeling is the act of taking several documents of the same type and adding tags around the content until all the documents are tagged in a way that fulfills the user’s needs. Document modeling can also encompass the theoretical; describing structures and features that are likely to be found in a document, but may not be seen in a set of sample documents.

Once a document type has been sufficiently modeled, a complete set of elements and attributes available for that specific document type are pulled together to form a schema. A schema, also known as a tag set, provides groupings of elements and attributes, and defines the relationships between them. Schemas may be standardized, either by issuing “guidance” for how they are to be implemented or by a more official process involving a standards organization, such as NISO. In the case of JATS, which will be discussed further below, standardization has taken place not at the level of the schema, but at a higher level of the *definitions* of the elements and attributes themselves (i.e. not at the level where the *relationships* between the elements and attributes are defined). JATS refers to this higher level as a Tag Suite. From the Tag Suite a number of different schemas, or tag sets, can be made. The documentation for each JATS Tag Set is referred to as a Tag Library, which allows users to browse a list of elements and attributes, and includes examples and definitions.

After becoming familiar with the topics discussed above, I was provided with several exercises to apply this knowledge to practice. In learning about XML and document modeling, I was given three recipes and asked to create a tag set that could be used to tag all of the recipes. This exercise highlighted how decisions made during the modeling process affect how well XML represents a document and in turn, how well XML functions when being used for information retrieval, processing, and analysis.

## JATS Standard

When the National Library of Medicine’s (NLM) National Center for Biotechnology Information (NCBI) created PubMed Central (PMC), a database of full-text biomedical journal articles, they required publishers to supply articles in either Standard Generalized Markup Language (SGML) or Extensible Markup Language (XML) along with any necessary images so that articles could be displayed at the PMC website. Some disadvantages from this model soon became apparent. PMC had to accommodate each type of publisher SGML or XML model coming in, and every time a new publisher began submitting content, the software had to be re-written to accommodate another model (Beck, 2011).

In 2001, the PMC workflow was modified so all publisher SGML or XML articles being submitted to the database were converted to a common XML structure, known as the NLM DTD. The NLM DTD went through several iterations from 2003 to late 2010. In 2006, the British Library and the U.S. Library of Congress announced their support formalizing the NLM DTD into a NISO standard. In 2011, the first version of JATS was approved by NISO as Z39.96x, JATS: Journal Article Tag Suite (Beck, 2011).

**Journal Article Tag Suite (JATS)**

**Figure 3. Journal Article Tag Suite** defines a set of XML elements and attributes for tagging journal articles and describes three article models. The tags available in these models are found in three Tag Sets: the Journal Archiving and Interchange, Journal Publishing, and Article Authoring.

The National Library of Medicine maintains three non-normative JATS schemas and other supporting documents which represent the NISO standard. The article models available for use are outlined in Figure 3. The Green model is used for archiving purposes, and includes the broadest set of elements and attributes. The archive PORTICO (<http://www.portico.org>) uses the green model. The Blue model is used by the journal publishing community it provides the most commonly used elements and attributes while disallowing the use of others. As represented in Figure 3, the Blue model is a subset of the Green model. The Orange model is meant as an authoring tool and is a highly restricted subset of the Blue model. Each model is available in three XML schema languages: DTD, W3C XML Schema, and RELAX NG.

## Supplementary Material

Supplementary materials are single information objects (such as a video file or excel workbook), or a physical container (such as a zip file with many items within, or a PDF file with embedded tables and figures). When publications were distributed in print only, supplementary material was considered either traditional or non-traditional. Traditional supplementary material was anything that, due to page limitations, would not fit in the text of a printed article. Non-traditional supplementary material included anything that could not be printed in the text of the article because of formatting issues, such as a movie file or computer program.

As online publishing began and grew, authors took the opportunity to add more additional content to their articles because the effort and cost was relatively low. The lack of recognized guidelines for dealing with this additional content soon became apparent. Publishers and archives made their own decisions about which types of information should be considered supplementary to an article, how to deliver supplementary information to users, how to make these materials discoverable by users, and how to preserve supplemental information.

The Journal of Cell Science imposes an 8000 word and 8 figure limit for articles, as well as a 5 MB supplementary material limit per article. The digital version of the journal is meant to be a close replica of the print version; to accomplish this, the publishers implemented a policy stating that supplementary material should only be comprised of “data that cannot be present in the main text” such as movies and large datasets. They believe that textual materials such as methods, results or discussions belong in the article itself, and so they do not publish supplementary text files (Journal of Cell Science, 2013).

In late 2010 the journal the Journal of Neuroscience implemented a policy that they would no longer allow authors to include supplementary material when they submitted new manuscripts. The journal also decided not to host supplemental material on its website. Instead, authors are allowed to include a footnote with a URL that points to supplemental material on a site which is supported and maintained by the author. None of the supplemental materials are reviewed or hosted by the journal (Journal of Neuroscience, 2013).

Even archivists and other non-publishers had differing ideas about tagging the supplementary material with XML. Some tagged the content as auxiliary information, while others considered it supporting information. JATS provided a body level element <supplementary-material>, but even this structure seemed inadequate to handle the increasingly complex load of figures, graphs, table, zip files, images and text being classified as supplementary to an article.

## NISO Recommendation

In 2009, NISO put together a committee to identify and address issues related to supplementary material and in 2013 the committee published a set of best practices concerning supplementary material. The [NISO Recommended Practices](http://www.niso.org/publications/rp/rp-15-2013) (referred to as the NISO RP) (http://www.niso.org/workrooms/supplemental) is divided into two parts: Part A provides recommendations on selecting, editing, managing and hosting supplementary material, and defines the responsibility of various parties to maintain discoverability of supplementary materials. Part B contains recommendations for providing adequate metadata and persistent identifiers for supplementary material, as well as the information required for preservation of the materials. Finally, the recommendation includes an example non-normative DTD, a tag library, and examples of articles tagged with the recommended elements and attributes.

The NISO RP challenged the traditional and non-traditional definitions of supplementary materials by creating a framework which identifies the level of importance of an information object to an article. Information objects could be considered either additional or integral to an article, depending on the importance designated by the author or publisher.

Supplementary Additional Material was defined as content that provides additional, relevant, and useful expansion of the article in the form of text, tables, figures, multimedia, or data, and that may aid any reader to achieve deeper understanding of the current work through added detail and context. The NISO RP recommends for best practices that while the author creates the content, the publisher would take responsibility to host it. Additional content may be so large that resources for editing it may not be available. For this reason, it is recommended that publishers indicate either in the material itself or in the article the level of editing the additional content has undergone. (NISO, 2013).

Supplementary Integral Material was defined as content essential for the full understanding of the work by the general scientist or reader in the journal’s discipline, but which is placed outside the article for technical, business, or logistical reasons. An example of supplementary integral information is a movie that the author regards as very important to the understanding of the article. NISO recommends that generally the publisher should maintain the responsibility for hosting and curating the content in the same manner as the article. They also recommend that any editing should be done at the same level as the core article (NISO, 2013).

# Preliminary Work

One of the best ways to become familiar with recommendations like the NISO RP, and to determine how well it works, is to use it. After a close reading of the NISO RP, I used the recommended elements and attributes to tag examples of supplemental materials from several articles. I then tagged the same examples using the JATS (Blue Model) DTD. What I found was there were many concepts addressed directly in the NISO RP that were not explicitly present in JATS. For example, NISO has elements that explicitly mark-up information related to preservation where JATS does not. My experience tagging the articles formed the basis of discussion among the group about whether the recommendations were able to represent the supplementary material adequately, whether they were easy to use, and how they compared to JATS.

Figure 5 contains an example of supplementary material that I will use to compare and contrast the JATS and NISO RP supplementary material XML structure, and in a later section it will be used to showcase the JATS extension.

**Figure 5.** Supplementary material (as seen in PMC) containing one additional dataset. On the left is a representation of the Supplementary Material section with one additional file in an online format. On the right is the additional file in Excel.Supplementary material (as seen in PMC) containing one additional dataset. On the left is a representation of the Supplementary Material section with one additional file in an online format. On the right is the additional file in Excel.

Meredith, B.K. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland, *BMC Genet.13*(21).

### JATS Supplementary Material Structure

In JATS, the root tag of an entire article is the element <article>. Each article has front matter represented by the element <front> consisting of journal and article metadata. Articles also have body matter represented by the element <body> containing the journal article, including text, figures, images, graphics and so on. A section of the body labeled “Supplementary Material” is found at the end of the article body. Because supplementary material is found within the body of the article, it is always associated with an article. The <supplementary-material> element can hold descriptive information such as labels, captions and titles, as well as <graphic>, <media> and <table-wrap>. Each of these is recursive and can appear within themselves. The supplementary material structure in JATS is noted in Figure 6, with a further explanatory chart in Appendix 5, Figure S1.

**Figure 6.** Example of JATS Supplementary Material tagging using the supplementary material from Figure 5.

<article xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:mml="http://www.w3.org/1998/Math/MathML" article-type="research-article">

<front>

<journal-meta> Journal metadata here</journal-meta>  
<article-meta> Article metadata here </article-meta>

</front>

<body> Journal article text here (background, results, conclusions, figures, images, etc.)

<sec sec-type="supplementary-material">

<title>Supplementary Material</title>  
<supplementary-material content-type="local-data" id="S1">

<caption>

<title>Additional file 1</title>  
<p><bold>Summary statistics for the cow dataset</bold>. This file contains summary statistics for the phenotypic data in the cow validation dataset mentioned in the manuscript. Summary statistics include the total number of phenotypic records (N), mean and standard deviation for each trait in the cows. Phenotypes in the cows are expressed as yield deviations on a PTA scale.</p>

</caption>  
<media xlink:href="1471-2156-13-21-S1.XLS" mimetype="application" mime-subtype="vnd.ms-excel">

<caption>

<p>Click here for file</p>

</caption>

</media>

</supplementary-material>

</sec>

</body>

<back> Acknowledgements and References here </back>

</article>

### NISO RP Supplementary Material Structure

The NISO RP was deliberately made to be a very recursive model. The root element for the supplementary material is <sup-mat-metadata> which wraps the entire supplementary material section. The element <sup-mat-metadata> wraps each information object inside the supplementary material. Unlike JATS in which the element <supplementary-material> is meant to be contained within the body of an article, the NISO RP model assumes the supplementary material is not associated with an article, and therefore must contain that type of information in the element <article-metadata>. The elements <sup-mat-metadata> and <sup-mat-object> include descriptive information, such as a summary or label, title and caption of an information object which are wrapped together in the element, <sup-mat-descriptive>. The element <sup-mat-object> might also contain further physical information, such as a file name, mime-type or an external link. These items of information are wrapped together in the element <sup-mat-physical>. While not shown in the example below, the element <sup-mat-object-metadata> is recursive and can contain itself <sup-mat-object-metadata>. An example of the NISO RP Supplementary Material structure is noted in Figure 7 below, with a further explanatory chart in Appendix 5, Figure S2.

**Figure 7.** Example of NISO RP Supplementary Material

<sup-mat-metadata dtd-version="1.00" sup-mat-type="integral">

<article-metadata>

<article-verification type="article\_citation">Meredith, B.K. (2012), Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland, BMC Genet. 13(21). </article-verification>

</article-metadata>  
<sup-mat-descriptive>

<summary>An XSL file for Additional file 1</summary>

</sup-mat-descriptive>

<sup-mat-object-metadata sup-mat-type="additional" relationship-of-objects="single">

<sup-mat-descriptive>

<sup-mat-identifier>10.1186/1471-2156-13-21.XSL001</sup-mat-identifier>  
<label>Click here for file</label>  
<titles>

<title>Additional file 1</title>

</titles>  
<caption><b>Summary statistics for the cow dataset</b>. This file contains summary statistics for the phenotypic data in the cow validation dataset mentioned in the manuscript. Summary statistics include the total number of phenotypic records (N), mean and standard deviation (σ) for each trait in the cows. Phenotypes in the cows are expressed as yield deviations on a PTA scale.</caption>

</sup-mat-descriptive>

<sup-mat-physical>

<file-metadata>

<filename>1471-2156-13-21-S1.XLS</filename>  
<format>XLS</format>  
<mime-type>application</mime-type>  
<mime-subtype>vnd.ms-excel</mime-subtype>  
<size>14 KB</size>

</file-metadata>

</sup-mat-physical>

</sup-mat-object-metadata>

</sup-mat-metadata>

# An Extension

**Figure 8**. Creating an Extension to Journal Article Tag Suite (JATS).

Creating an Extension to Journal Article Tag Suite (JATS)The purpose of this project was to develop an extension to JATS which when used in conjunction with JATS would fully support the NISO RP’s definition of supplemental additional and supplemental integral information in journal articles. Figure 8 illustrates how JATS, the NISO RP and the professional experiences of project leaders, Jeff Beck and Kimberly A. Tryka brought about the creation of a JATS Extension which will be referred to in this paper as EXT.

There were several ongoing discussions that defined the creation of the extension. The issues of supplementary material such as the movement from a traditional vs. nontraditional definition to the material’s level of importance to an article were interwoven in all of our discussions. As we worked through the process of creating and discarding, and creating again, an extension, we discussed how the extension might work from a document modeling standpoint. These discussions allowed us to create hierarchical relationships between elements to best represent a file and its contents. The element relationships changed based on our attempts to use the extension on current supplementary materials. Perhaps the longest running discussion during this project was based on the objectives of the NISO RP, and whether or not our model in its various states supported the objectives as best as possible.

We were mindful that the NISO RP represented a standalone XML structure, one in which the article for which the material was supplemental did not exist. In contrast, supplementary material found in a JATS environment would always be in relationship to an article. Because the JATS environment already included a complex article structure, we knew the NISO RP example XML structure could not be simply overlaid upon the JATS structure. We did several mapping projects between the NISO RP structure and the JATS structure in order to find similar structural arrangements and contemplate areas of deficiencies.

As the extension came together, we sought out interactions with several key members of the NISO RP working group. Kathy Kwan from NLM’s National Center for Biotechnology Information (NCBI) provided insight into the file structure and preservation issues that surround supplementary material. Deborah Lapeyre from Mulberry Technologies provided background information on Standardized General Markup Languages, and listened to our description of the extension. Alexander Schwarzman of The Optical Society discussed the nuances of the definitions of supplementary material.

For a further breakdown of the timeline for discussions, technical training, document modeling work, mapping between structures, and other work completed for this project, see Appendix 4.

## The JATS Extension (EXT)

In order to create an extension of JATS, we had to make some decisions early on which would help set our path. In some areas we completely re-worked the NISO RP structures to suit our purposes, and in other areas, we adopted a close approximation to the NISO RP structure.

Our first decision was to consider that while the NISO RP was created as a standalone model for supplementary material, we would consider the supplementary material section in JATS to exist within the body element of an actual article, and not outside or away from the article. Therefore we would not repeat information in the supplementary material element that could be associated with the article.

Second we decided to work within the current JATS framework as much as possible by preferring JATS elements or structures over those which hold the same or similar meaning as NISO RP elements. We would add NISO RP elements or models, and deprecate JATS elements only when necessary. In addition any NISO RP elements or models added to the extension could undergo a name change if we felt it would help people understand the purpose of an element or attribute.

Our final decision was to open up the definition of supplementary materials and consider any item not in the core text of the article to be supplementary. This brought elements found within the JATS body element such as <fig>, <graphic>, <media> and <table-wrap> up for scrutiny, along with <supplementary-material>. We were able to create a leaner structure in which <fig> and <supplementary-material> had the same models, and <graphic> and <media> were deprecated.

## Major Revisions

Revisions were made to the NISO RP in order to create an EXT that supported the NISO RP and worked within the existing structure of the JATS. Two major revisions to NISO RP elements, <sup-mat-metadata> and <sup-mat-object-metadata> are discussed below. Other minor revisions to the NISO RP can be found in Appendix 5. The concept for expressing these revisions came from documentation written by Kimberly A. Tryka to support the Independent Validation and Verification of the extension by Mulberry Technologies, and are reiterated below with additions. Several example articles were tagged with the EXT and can be linked to from Appendix 7.

### Revisions to the NISO RP

**Parent element:** <sup-mat-metadata>

* **Definition:** Wrapper for one or more supplemental objects that relate to an article or, more correctly, the metadata describing those supplemental objects. This element is intended to be the root element of the supplemental material XML file (NISO, 2013).
* **EXT:** There is no equivalent element. Therefore, there is no overall wrapper holding all supplemental material objects. The child elements of <sup-mat-metadata> could be adequately expressed in other locations.
* **Child Elements:**
  + The content found in <date> and <article-metadata> would be adequately expressed in JATS element, <front>.
  + The content found in <sup-mat-descriptive> could be found in JATS element, <front>. A summary of the supplemental material is now found in a new element <ext:summary> as a child of JATS element, <front>.
  + The element <sup-mat-object-metadata> becomes the highest level for supplementary material. However, this element is separated into EXT elements, <supplementary-material> or <fig> elements, as is outlined below.

**Parent element:** <sup-mat-object-metadata>

* **Definition:** Container element for the metadata concerning one supplemental object (NISO, 2013).
* **EXT:** Separated into two elements, <supplementary-material> or <fig> elements
* **Child Elements:**
  + The content found in <article-component-metadata> is adequately described in JATS element, <front>.
  + The child element <sup-mat-descriptive> wrapped around elements such as <sup-mat-identifier>, <version>, <label>, <titles>, <caption>, <history>, <publisher>, <copyright>, <preservation>. There is no equivalent element in the EXT. Instead the children of <sup-mat-descriptive> can be found without a wrapper under EXT elements <supplementary-material> and <fig>.
  + The child element <sup-mat-object-metadata> is no longer retained.
    - **EXT <ext:resource>:** Any recursion under EXT elements <supplementary-material> and <fig> in which the information object is addressable in a file system is wrapped in the new JATS EXT element, <ext:resource>.
    - **EXT <ext:item>:** Any recursion under the EXT element <supplementary-material>, <fig>, and <ext:resource> in which the information object is not addressable in a file system, but is still able to be references (by a label, title or other ID) is wrapped in the new EXT element, <ext:item>.
  + The child element <sup-mat-physical> is no longer retained, as well as its child element <file-metadata> which acted as a wrapper for file information. Instead, any file information is found under the new EXT element, <ext:resource-meta> which is a child of new EXT element, <ext:resource>.

### Revisions to the JATS Journal Publishing Model

**Parent elements:** <fig> and <supplementary-material>

* **<fig> definition:** Block of graphic or textual material that is identified as a figure, usually bearing a caption and a label such as “Figure 3.” or “Figure”.
* **<supplementary-material> definition:** Container element for description of and pointer to additional data files that contain information directly supportive of the document (for example, audio clip, applet, raw research data).
* **EXT:** Both elements are continued as children of parent element <body>.
* **Previous Child Elements:** 
  + **<graphic>:** This element is deprecated in the EXT.
  + **<media>:** This element is deprecated in the EXT.
* **New Child Elements:**
  + **EXT <ext:resource>:** Any recursion under EXT elements <supplementary-material> and <fig> in which the information object is addressable in a file system is wrapped in the new EXT element, <ext:resource>.
    - File information is found under the new EXT element, <ext:resource-meta> which is a child of new EXT element, <ext:resource>.
  + **EXT <ext:item>:** Any recursion under the EXT element <supplementary-material>, <fig>, and <ext:resource> in which the information object is not addressable in a file system, but is still able to be references (by a label, title or other ID) is wrapped in the new EXT element, <ext:item>.

### Tag Libraries

Tag libraries are browse-able lists of elements and attributes, and include definitions and structural information, such as descriptions of sibling and child elements. The respective tag libraries serve as excellent resources in order to understand existing structures and where revisions have been made.

* The NISO RP Tag Library (<http://supplemental.niso.org/tag-library/index.html>)
* JATS (Journal Publishing Tag Set) Tag Library: http://jats.nlm.nih.gov/publishing/tag-library/1.0/
* EXT Tag Library: <http://www.ncbi.nlm.nih.gov/staff/beck/sme/el-fig.html>.

### EXT Supplementary Material Structure

The EXT emerged as a result of both the major and minor revisions. The extension is not a standalone document and is intended to be used with the JATS Journal Publishing tag set (Blue Model) and the JATS non-normative DTD. An example of the supplementary material structure in the EXT is noted in Figure 9, with a further explanatory chart in Appendix 6, Figure S3. In addition, several articles tagged in JATS EXT XML are linked to in Appendix 7.

**Figure 9.** Example of EXT Supplementary Material.

<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE article SYSTEM "…\NISO-supmat-extension-rec0.dtd">  
<article dtd-version="1.0" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:mml="http://www.w3.org/1998/Math/MathML" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >

<front>

<journal-meta></journal-meta>  
<article-meta></article-meta>

</front>

<body> Journal article text here (background, results, conclusions, figures, images, etc.)

<sec sec-type="supplementary-material">

<title>Supplementary Material</title>  
<supplementary-material content-type="local-data" id="S1" xlink:href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3361482/bin/1471-2156-13-21-S1.XLS">

<caption>

<title>Additional file 1</title>  
<p><bold>Summary statistics for the cow dataset</bold>. This file contains summary statistics for the phenotypic data in the cow validation dataset mentioned in the manuscript. Summary statistics include the total number of phenotypic records (N), mean and standard deviation for each trait in the cows. Phenotypes in the cows are expressed as yield deviations on a PTA scale.</p>

</caption>  
<ext:resource xlink:href="1471-2156-13-21-S1.XLS" mimetype="application" mime-subtype="vnd.ms-excel">

<ext:resource-meta>

<ext:filename>1471-2156-13-21-S1.XLS</ext:filename>  
<ext:format>XLS</ext:format>  
<size units="KB">14 </size>

</ext:resource-meta>  
<ext:item>

<caption>

<title>Additional file 3- summary statistics for the cow dataset</title>  
<p>Summary statistics include the total number of phenotypic records (N), mean and standard deviation for each trait in the cows.</p>  
<p>Phenotypes in the cows are express as yield deviations on a PTA scale</p>

</caption>

</ext:item>

</ext:resource>

</supplementary-material>

</sec>

</body>  
<back> Acknowledgements and References here </back>

</article>

# Recommendations

The EXT was sent to Mulberry Technologies, Inc. as part of an independent validation and verification (IV&V) review. Their comments provide us with the opportunity to reconsider the structures we’ve created or confirm the soundness of our design. A preliminary comment from their report provides some insight as to their recommendation for our work.

“…We consider this a successful exercise, because the tag set created would achieve the stated goals, in many cases very well, and far more important, it reveals several issues that must be resolved before a production version of this enhanced tagging can be developed….An experiment that points out this many problems is, by definitions, a success (Mulberry Technologies, 2013).”

One recommendation that I would suggest for the future is to revisit an idea that had previously been set aside. As a thought exercise, we considered the possibility of deprecating the <supplementary-material> element, and instead allowing any non-textual object in an article (whether a figure, image, table, or supplementary material) to have an attribute explaining the level of importance of the object to the article. On which element this attribute would be placed has yet to be determined. This practice would permit publishers, authors, and others to move away from the concept of traditional versus nontraditional supplemental information, and move toward better qualifying information in relation to an article, regardless of its print status.

# Acknowledgements

I wish to acknowledge the project managers, Jeff Beck, NCBI Technical Information Specialist, and Kimberly A. Tryka, NCBI Staff Scientist for the guidance, support and training they provided to me during this project. I would also like to acknowledge Kathel Dunn, NLM’s Associate Fellowship Coordinator for her support and oversight of this project. I am thankful for the members of NISO’s Technical Working Group, Kathy Kwan, Deborah Lapeyre, and Alexander Schwarzman for sharing their knowledge of supplementary material and XML structures. Finally, I would like to thank the National Library of Medicine, Dr. Lindberg, Betsy Humphreys, Joyce Backus and Dianne Babski for providing me with the opportunity to learn from and work with the National Library of Medicine.

# References

Beck, J. (2011) NISO Z39.96 The Journal Article Tag Suite (JATS): What Happened to the NLM DTDs? *Journal of Electronic Publishing, 14*(1). doi: <http://dx.doi.org/10.3998/3336451.0014.106>

Coombs, J.H., Renear, A.H., DeRose, S.J. Markup systems and the future of scholarly text processing. (1987). *Communications of the AMC,* *30*(11): 933-47.

Journal of Cell Science. (2013). Supplementary material. Retrieved from: http://jcs.biologists.org/site/author/supplementary\_material.xhtml

Journal of Neuroscience. (2013). Supplemental material. Retrieved from: http://www.jneurosci.org/site/misc/ifa\_supplemental.xhtml

National Information Standards Organization. (2013) Recommended Practices for Online Supplemental Journal Article Material. 2013 Jan. Retrieved from: http://www.niso.org/workrooms/supplemental

Renear, A. (2004). Text Encoding. In Schreibman, S. (Eds.), *A Companion to Digital Humanities*. Oxford: Blackwell. Retrieved from http://www.digitalhumanities.org/companion/

TEI Consortium. (2013). A Gentle Introduction to XML. Retrieved from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>

Wikipedia. (2013). XML Schema Language. Retrieved from: http://en.wikipedia.org/wiki/XML\_schema\_languages

# Appendix 1: Glossary of Terms

| **Term** | **Definition\* *\*all definitions are made in reference to JATS, and or tagging journal articles.*** |
| --- | --- |
| Ancestry | Refers to the concept of elements having hierarchal relationships to one another.  Ex. <article> has children <article-meta> and <paragraph>. <article-meta> has a child <article-id>. |
| Article Model | * A defined set of tags used as a model to describe the structure and content of an article. * The Journal Article Tag Suite (JATS) provides three article models from which a user may choose when describing the structure and content of an article.   + Journal Archiving and Interchange- the most permissive of the tag sets   + Journal Publishing- a moderately prescriptive tag set   + Article Authoring- the most prescriptive of the tag sets * JATS website: <http://jats.nlm.nih.gov/> |
| Attribute | * Modifier of elements found in the opening tag. * Ex. <article language= “English”> text </article> where language is the attribute and English is the value of the attribute. * Attributes often provide information that is not explicitly part of the text of an article.   + Ex. While the language of an article is surmised visually, a language attribute specifically identifies the language within the tags as English. * Attribute values can be based on a controlled vocabulary. However there is no way to enforce compliance with the vocabulary. |
| Document Modeling | Refers to the creation or modification of XML and/or a Document Type Definition through the process of examining existing article structures and/or identifying theoretical structures for articles not found in current practice. |
| Document Type Definition (DTD) | * When tagging a specific type of document in XML, the DTD defines: * the elements and attributes which are permissible for use * the structure in which elements and attributes may be used * what is allowed inside the elements (i.e. #PCDATA) * A DTD may be declared in an XML document, or as an external reference   + Ex. of an externally referenced DTD: <!DOCTYPE article SYSTEM "../dtd/NISO-supmat-extension-rec0.dtd"> * DTDs are one of three main XML Schema Languages (see definition provided). * DTDs are not necessary when using XML. However, DTD provide several advantages:   + DTDs allow XML files to carry descriptions of their intended structure   + Standard DTDs allow for the interchange of articles between independent groups.   + Using a DTD allows for the verification and validation of internally or externally tagged articles. |
| Element | * Represent an object or text in a document. * Surrounded by angled brackets and closes with a forward slash.   + Ex. <element-name> article text here </element-name> * The opening and closing instances of an element are often referred to as “tags” * Must be lowercase, one word or compound terms * Should be concise |
| NISO | * National Information Standards Organization is a non-profit association accredited by the American National Standards Institute (ANSI). * NISO identifies, develops, maintains and publishes technical standards to manage information. * NISO website: <http://www.niso.org/home/> |
| Parent/Child | Refers to the concept of elements having hierarchal relationships to one another.  Ex. <article> has children <article-meta> and <paragraph>. <article-meta> has a child <article-id>. |
| Root element | Encloses all other elements in a document. Each article has exactly one root element. Ex. <article> |
| Siblings | Refers to the concept of elements being equally related to one another. Ex. <article-meta> and <paragraph> are siblings under the parent element <article>. |
| Tag | The opening and closing instances of an element are often referred to as “tags” |
| Tag Set | * Refers to an article model built from the elements and attributes in a Tag Suite (see definition). * The Journal Article Tag Suite (JATS) provides three article models from which a user may choose when describing the structure and content of an article.   + Journal Archiving and Interchange- the most permissive of the tag sets   + Journal Publishing- a moderately prescriptive tag set   + Article Authoring- the most prescriptive of the tag sets * JATS website: <http://jats.nlm.nih.gov/> |
| Tag Suite | The complete set of elements and attributes that have been defined to tag journal articles. Tag Suites are not entirely common in the XML world. They are considered a structure or level above the Tag Set, in which several tag sets are pulled together to form one Tag Suite. JATS is considered a Tag Suite. |
| Tagging | The process of adding elements and attributes to an article. |
| Tag Library | A browse-able list of elements and attributes |
| XML | * e**X**tensible **M**arkup **L**anguage * Defines a set of rules for encoding documents in a format that is both human-readable and machine-readable * Focuses on describing the structure and meaning of a document   Taken from: XML. Wikipedia. Retrieved on 6/26/2013 from http://en.wikipedia.org/wiki/XML |
| XML schema language | * Used to express a set of rules to which an XML document must conform in order to be considered 'valid' according to that schema. * There are three main XML schema languages:   + WXS (also known as “XML Schema” and XSD)     - published in 2001 by W3C     - written in XML     - designed with the intent that determination of a document's validity would produce a collection of information adhering to specific data types.   + Document Type Definition (DTD)     - Used with SGML before XML was created     - Includes element declarations and attribute list declarations   + RELAX NG     - **RE**gular **LA**nguage for **X**ML **N**ext **G**eneration     - written in XML   Taken from: XML schema Languages. Wikipedia. Retrieved on 6/26/2013 from http://en.wikipedia.org/wiki/XML\_schema\_languages |

# Appendix 2. Example of XML (Recipe)

<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE recipe SYSTEM "newrecipe-kat.dtd">  
<recipe>

<header>

<title>Original NESTLÉ® TOLL HOUSE® Chocolate Chip Cookies </title>  
<author> NESTLÉ TOLL HOUSE</author>  
<image file="http://www.meals.com/imagesrecipes/18476lrg.jpg" alt-text="Toll House cookie picture"/>  
<p> This famous classic American cookie is a treat no matter what the age or occasion. Enjoy it with a glass of cold milk.</p>  
<p>http://www.verybestbaking.com/recipes/18476/Original-NESTL%C3%89-TOLL-HOUSE-Chocolate-Chip-Cookies/detail.aspx</p>

</header>  
<ingredient-list>

<title>Ingredients</title>

<ingredient amount="2 1/4" units="C"><name>all-purpose flour</name></ingredient>  
<ingredient units="tsp" amount="1"><name>baking soda</name></ingredient>  
<ingredient amount="1" units="tsp"> <name>salt</name></ingredient>  
<ingredient units="C" amount="1"><name>butter, softened</name><comment>(2 sticks)</comment></ingredient>  
<ingredient amount="3/4" units="C"><name>granulated sugar</name></ingredient>  
<ingredient amount="3/4" units="C"><name>packed brown sugar</name></ingredient>  
<ingredient units="tsp" amount="1"><name>vanilla extract</name></ingredient>  
<ingredient amount="2"><name>large eggs</name></ingredient>  
<ingredient amount="2" units="C" ><name>NESTLÉ® TOLL HOUSE® Semi-Sweet Chocolate Morsels</name><comment>(12-oz. pkg.) </comment></ingredient>  
<ingredient units="C" amount="1"><name>chopped nuts</name></ingredient>

</ingredient-list>

<instructions>

<title>Directions</title>  
<steps> <step>PREHEAT oven to 375° F.</step>

<step>COMBINE flour, baking soda and salt in small bowl. Beat butter, granulated sugar, brown sugar and vanilla extract in large mixer bowl until creamy. Add eggs, one at a time, beating well after each addition. Gradually beat in flour mixture. Stir in morsels and nuts. Drop by rounded tablespoon onto ungreased baking sheets.</step>

<step>BAKE for 9 to 11 minutes or until golden brown. Cool on baking sheets for 2 minutes; remove to wire racks to cool completely.</step>

</steps>

</instructions>  
<variation>

<title >PAN COOKIE VARIATION: </title> <p>Grease 15 x 10-inch jelly-roll pan. Prepare dough as above. Spread into prepared pan. Bake for 20 to 25 minutes or until golden brown. Cool in pan on wire rack. Makes 4 dozen bars.</p>

</variation>

<variation>

<title> SLICE AND BAKE COOKIE VARIATION:</title>

<p>PREPARE dough as above. Divide in half; wrap in waxed paper. Refrigerate for 1 hour or until firm. Shape each half into 15-inch log; wrap in wax paper. Refrigerate for 30 minutes.\* Preheat oven to 375° F. Cut into 1/2-inch-thick slices; place on ungreased baking sheets. Bake for 8 to 10 minutes or until golden brown. Cool on baking sheets for 2 minutes; remove to wire racks to cool completely. Makes about 5 dozen cookies.</p></variation>

</recipe>

# Appendix 3. Example of DTD (Recipe)

<?xml version="1.0" encoding="UTF-8"?>  
<!ELEMENT recipe (header, ingredient-list, instructions, variation\*) >  
<!ELEMENT header (title, subtitle?, author\*, image?, p\* ) >  
<!ELEMENT subtitle (#PCDATA) >  
<!ELEMENT author (#PCDATA) >  
<!ELEMENT image EMPTY>  
<!ATTLIST image

File CDATA #REQUIRED  
alt-text CDATA #REQUIRED>

<!ELEMENT ingredient-list (title?, ingredient+) >  
<!ELEMENT title (#PCDATA) >  
<!ELEMENT ingredient (name, preparation?, comment?) >  
<!ATTLIST ingredient

amount CDATA #IMPLIED  
units (C | T | tsp) #IMPLIED>

<!ELEMENT name (#PCDATA) >  
<!ELEMENT preparation (#PCDATA)>  
<!ELEMENT comment (#PCDATA) >  
<!ELEMENT p (#PCDATA | bold)\* >  
<!ELEMENT bold (#PCDATA) >  
<!ELEMENT instructions (title?, p\*, steps?) >  
<!ELEMENT steps (step+) >  
<!ELEMENT step (#PCDATA) >  
<!ATTLIST step

label CDATA #IMPLIED>

<!ELEMENT variation (title, p\*, steps?) >

# Appendix 4: Gantt chart of Project Duration

# Appendix 5. Schematic representation of XML Structures

**Figure S1.** Schematic representation of **Journal Article Tag Suite** **(JATS)** supplementary material model. *Please note that not all elements and attributes are included in this example.*

OR

OR

**Figure S2.** Schematic representation of **NISO Recommended Practices for Online Supplemental Journal Article Materials** for both integral and additional supplementary material. Elements of likeness are represented in similar colors to highlight the recursive use of elements in this model. *Please note that not all elements and attributes are included in this example.*

**Figure S2.** Schematic representation of **extension to Journal Article Tag Suite for supplemental integral materials** based on recommendations. *Please note that not all elements and attributes are included in this example.*

# 

**Figure S3.** Schematic representation of **extension to Journal Article Tag Suite for supplemental additional materials** based on recommendations. *Please note that not all elements and attributes are included in this example.*

# Appendix 6. Minor Revisions between NISO RP and JATS

<accessibility-long-desc> is replaced by the JATS element <long-desc>

<contrib-group> as modeled in NISO is replaced by the JATS model for <contrib-group>

<titles> and <caption> as modeled in NISO is replaced by the JATS model for <caption>

<b> and other face-markup elements are replaced by their JATS equivalents

<content-descriptor> is replaced by JATS @content-type

<copyright>, <license> and their children are replaced by the JATS <permissions> model

<creation-date>, <date>, <pub-date> and their children are replaced by the JATS <date> model (distinctions between dates can be made by using @date-type)

<ext-link> as modeled in NISO is replaced by the JATS model for <ext-link>

<graphic> as modeled in NISO is replaced by the JATS model for <graphic>

<history> as modeled in NISO is replaced by the JATS model for <history>

<mime-type> and <mime-subtype> are replaced by the equivalent attributes @mime-type and @mime-subtype from the JATS model

<open-access> has not been retained – seems to be adequately covered by the JATS model for <permissions>

<publisher> and its children as modeled in NISO is replace by JATS model for <publisher>

<preservation> has been renamed <preservation-commitment> and the model has been added to the JATS model for <supplementary-material> and <fig>

<provenance> has been renamed <original-sources> and the model has been added to the JATS model for <supplementary-material> and <fig>

<subject-descriptor-group> as modeled in NISO is replaced by the JATS model for <kwd-group>

@relationship-of-objects has been dropped. If the value was “alternatives” the <alternatives> element should be used. The JATS attribute @content-type can be used if it is necessary to make explicit the concepts of “logical set” or “unrelated”.

# Appendix 7. Listing of Examples showing JATS EXT markup

**NISO examples in JATS EXT:**

| NISO sample article | Comments about file | Name of file tagged using JATS-SME |
| --- | --- | --- |
| Pourquié, O. (2011), Vertebrate segmentation: From cyclic gene networks to scoliosis, *Cell 145*(5), 650–663,  <http://dx.doi.org/10.1016/j.cell.2011.05.011> |  | Scoliosis-ext.xml |
| Garland, E., A. W. Goldizen, M. L. Rekdahl, R. Constantine, C. Garrigue, N. D. Hauser, M. M. Poole, J. Robbins, and M. J. Noad (2011), Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale, *Current Biology 21*(8), 687–691,  <http://www.sciencedirect.com/science/article/pii/S0960982211002910> |  | Whale-ext.xml |
| Huffard, C., R. L. Caldwell, and F. Boneka (2010), Male-male and male-female aggression may influence mating associations in wild octopuses (*Abdopus aculeatus*), *J. Comp. Psychol. 124*(1), 38–46,  <http://dx.doi.org/10.1037/a0017230> |  | Octopus-ext.xml |
| Micalizio, S., A. Godone, C. Calosso, F. Levi, C. Affolderbach, and F. Gruet (2012), Pulsed optically pumped rubidium clock with high frequency-stability performance, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 59*(3),  <http://dx.doi.org/10.1109/TUFFC.2012.2215> | Examples of <fig>/<alternates>/<ext:resource> | Clock-ext.xml |
| Yu J., et al. (2005), The genomes of *Oryza sativa*: A history of duplications, *PLoS Biol 3*(2), e38,  <http://dx.doi.org/10.1371/journal.pbio.0030038> | Example of zip/directory/file structure | Duplications-ext.xml |
| Marfan syndrome, in *Genes and Disease* [Internet], NCBI,  <http://www.ncbi.nlm.nih.gov/books/NBK22203> | Example of <fig>/<alternates>/<ext:resource> | Book-ext.xml |

**Random samples in JATS EXT:**

| Random sample article | Comments about file | Name of file tagged using JATS-SME |
| --- | --- | --- |
| Paul Welsh et al. (2010), Unraveling the Directional Link between Adiposity and Inflammation: A Bidirectional Mendelian Randomization Approach, J Clin Endocrinol Metab. 2010 January; 95(1): 93–99,  <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805500/> |  | Adiposity-ext.xml |
| Aune, U.L., Ruiz, L., and Kajimura, S. (2013), Isolation and differentiation of stromal vascular cells to beige/brite, *Journal of Visualized Experiments*. (73),<http://www.ncbi.nlm.nih.gov/pubmed/23568137>, <http://www.jove.com/video/50191/> | Contains 1 video which would be supplemental integral in the NISO Recommendation.  Uses <fig> with <ext:resource> | Isolation-ext.xml |
| Kaspi, Y., Showman, A.P., Hubbard, W.B., Aharonson, O. and Helled, R. (2013). Atmospheric confinement of jet streams on Uranus and Neptune. *Nature. (497),* pp. 344-347, <http://www.nature.com/nature/journal/v497/n7449/full/nature12131.html> | Only described at level of PDF (caption offers good description) – really should have been an appendix or two | Jet-ext.xml |
| Robert A. Sanford et al. (2012) Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *PNAS. 109(48)*. pp. 19709-19714, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3511753/> |  | Nitrous-ext.xml |
| Huang Liusheng et al. (2010). A modified method for determination of lumefantrine in human plasma by HPLC-UV and combination of protein precipitation and solid-phase extraction: application to a pharmacokinetic study. *Anal Chem Insights. (5)* pp. 15-23*,* <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2865164/> |  | Plasma-ext.xml |
| No current citation – forthcoming chapter on dbGaP from NCBI Handbook | Use of <original-sources> | Gap-ext.xml |
| <http://www.sciencemag.org/content/340/6138/1314.full> | Use of <history> | Redox-ext.xml |