

Extending JATS to include the NISO/NFAIS Recommended Practices for Online Supplemental Journal Article Materials

Karen Gutzman¹ and Kimberly A. Tryka².

¹ Bernard Becker Medical Library, Washington University School of Medicine, St. Louis, MO

² National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD

This paper discusses our experience of creating an extension for JATS that incorporates the NISO "Recommended Practices for Online Supplemental Journal Article Materials" (NISO RP-2013). We will discuss our analysis of the recommendations and our comparison of the recommendations with JATS, as well as our thrashing over language and terminology associated with supplementary materials and our eventual creation of the extension. The extension is not part of the official JATS specification; it is a local extension that will be made publicly available for community use and discussion.

Introduction

Following a survey about how publishers work with supplementary materials (conducted by Alexander (Sasha) Schwarzman, then at the American Geophysical Union), the National Information Standards Organization (NISO) and the National Federation of Advanced Information Systems (NFAIS) formed a working group to identify and address issues related to supplementary material. The results of that committee's work were published as a set of recommended best practices in January 2013, [NISO RP-15-2013, Recommended Practices for Online Supplemental Journal Article Materials](#), which we will refer to as the "NISO RP" for the remainder of this paper.

While it was too soon for PubMed Central (PMC) to start seeing these recommended practices being used in the content they were receiving, in the Spring of 2013 Jeffrey Beck proposed a project to the National Library of Medicine's (NLM) [Associate Fellowship Program](#) that aimed to create an extension of JATS that could absorb the information present in the NISO RP and that could be shared amongst the community for comment. The lead author of this paper, Karen Gutzman, was a NLM Fellow and decided to work on the project, advised by Jeffrey Beck and Kimberly Tryka. This paper is a result of that work.

In the end, while we succeeded in creating an extension that is available for public comment, we were left with doubts. In particular, although we managed to include the information found in the NISO RP, we found that we, even as creators, had trouble deploying the extension in a consistent manner.

Eventually we came to the conclusion that there was a deeper problem to be explored revolving around the concept of “supplementary-ness”. In practice, we find that what makes something supplementary, say a figure, is a policy decision (and thus varies between publishers and indeed between individual publications) rather than any difference in content. We have dubbed this way of interacting with additional materials “supplementary material exceptionalism” and we will speak about why ending this way of thinking about supplementary materials could offer a solution to the problems we saw in our extension.

Part I - The NISO RP and Creating Our JATS Extension

Before beginning work on the JATS extension we spent time trying to get a good feel for what the recommended practices were and how they had been implemented in the [non-normative DTD](#). While a complete summary of the recommended practices is beyond the scope of this paper, we will briefly describe some portions of the NISO RP that we feel are important for the discussion that follows. It should be noted that although the non-normative DTD (published as part of the [Supporting Documentation](#) for the NISO RP) is only one possible implementation of the recommended practices, we referred to it extensively during this project and will do so in this paper. Because the sample files included in the Supporting Documentation are marked up using the non-normative DTD it is hard not to use it as a touchstone, regardless of the “non-normative” label.

Content Defined by the NISO RP

In describing supplementary material, the NISO RP defines a series of recommended and optional metadata values (found in Table 1: Quick Reference to Supplemental Journal Article Material Metadata Recommendations). We have pulled out the values that are listed as required and summarized them in this document’s Table 1.

Table 1 Required content for Supplementary Material from the NISO-RP.

Hyperlinks will take you to the appropriate page from the NISO RP Tag Library for the non-normative DTD which gives a more detailed description of the element or attribute.

Concept	Definition	How this appears in non-normative DTD
Article Metadata	Information about the article in which supplementary material resides	<article-metadata> (wrapper element)
Article Identifier	An identifier, likely a DOI, but could contain any other identifier for the article	<article-metadata>/<article-identifier>
Supplemental Material Descriptive Metadata	Descriptive data either related to all the supplemental material, or just a particular piece of supplemental material (dependent on context)	<sup-mat-descriptive> (wrapper element)

Supplemental Material Identifier	An identifier, possibly a DOI or something domain specific	<sup-mat-descriptive> /<sup-mat-identifier>
Supplemental Material Object Metadata	The metadata related to a particular piece of supplementary material	<sup-mat-object-metadata> (wrapper element, recursive)
Relationship of Supplemental Material to the Article	Notes whether the material is necessary to understand the content of the article or is additional information	@sup-mat-type (values are “integral”, “additional” and “integral-and-additional”)
Supplemental Material Physical Metadata	Information about the file such as format, mime-type, size, fixity, etc.	<sup-mat-object-metadata>/<sup-mat-physical> (wrapper)
Supplemental Material File Formats	File format, preferably from a public format registry	<sup-mat-physical>/<file-metadata>/<format>

This is a minimal set of required information and it is focused on making sure that the supplementary materials have an identity, reference their parent (article), state their level of importance, and announce their format.

Table 1 in the NISO RP also lists the following optional content:

As part of Supplemental Material Object Metadata

- Supplemental Material Label
- Supplemental Material Title
- Supplemental Material Caption
- Supplemental Material Contributors
- Supplemental Material History
- Supplemental Material Subject Descriptor
- Supplemental Material Summary
- Supplemental Material Publisher
- Supplemental Material License
- Supplemental Material Provenance
- Relationship between Supplemental Material Objects
- Supplemental Material Version Number
- Original
- Article Component Metadata

As part of Supplemental Material Physical Metadata

- Supplemental Material Preservation
- Supplemental Material File Validity
- Supplemental Material Data Fixity
- Creating Application

- Rendering Application

In many ways these are the interesting parts of the recommendation; they flesh out the description of the supplementary material in such a way that you can get a very good idea of what the supplementary material contains as more of the optional data is added. For example, if the piece of supplementary material is a PDF, using only the minimum required information you will know the identifier of the PDF, the identifier of the article it is associated with, whether the information in the PDF is integral, additional or possibly both, and the fact that it is a PDF. By using the optional metadata you could further explain that this PDF contains 5 figures, 3 tables and an extended Methods section (using Supplemental Material Description), what the labels, titles, and captions of those figures and tables are (using Supplemental Material Label/Title/Caption), that Table S2 has been updated since the original publication (using Supplemental Material History), and that one of the figures was not made by any of the authors of the paper (using Supplemental Material Contributors).

As it is often much easier to “see” hierarchical data structures than to gain an understanding of them from a wordy description, we present a portion of a sample tagged file from the NISO RP Supporting Documentation in Box 1.

Box 1A portion of a tagged sample from the Supporting Documents for the NISO RP.

The content of the second <sup-mat-object-metadata> element has been removed to keep the example short. The full sample can be found at http://supplemental.niso.org/tagged-samples/10.1016_j.cell.2011.05.011-sup-mat-metadata-min.xml.

```
<sup-mat-metadata dtd-version="1.00" sup-mat-type="integral">
  <article-metadata>
    <article-identifier id-type="doi">
10.1016/j.cell.2011.05.011</article-identifier>
    </article-metadata>
    <sup-mat-descriptive>
      <summary>An MP4 file with Movie S1 and a MOV file with Movie
S2</summary>
    </sup-mat-descriptive>
    <sup-mat-object-metadata sup-mat-type="integral">
      <sup-mat-descriptive>
        <label>Movie S1</label>
        <titles>
          <title>Clock and Wavefront Model for Vertebrate
Segmentation</title>
        </titles>
      </sup-mat-descriptive>
      <sup-mat-physical>
        <file-metadata>
          <filename>PIIS009286741100537X.mmc1.mp4</filename>
          <format>MP4</format>
        </file-metadata>
        <ext-link href="http://download.cell.com/mmc/journals/0092-
8674/PIIS009286741100537X.mmc1.mp4"/>
      </sup-mat-physical>
    </sup-mat-object-metadata>
  </sup-mat-metadata>
```

```
</sup-mat-object-metadata>  
<sup-mat-object-metadata>Info for Movie S2</sup-mat-object-metadata>  
</sup-mat-metadata>
```

(Note that although the Supplemental Material Identifier is listed as required in the NISO RP, the non-normative DTD does not make the corresponding element `<sup-mat-identifier>` required, and thus it is not present in this particular tagged example.)

Mechanics of the Extension

We created the extension by making a new top-level DTD `NISO-supmat-extension-rec0.dtd`, which in turn calls the standard JATS DTD files for Journal Publishing and the following custom files for the changes/additions of the extension:

- `NISO-supmatcustom-classes.ent`
- `NISO-supmatcustom-elements.ent`
- `NISO-supmatcustom-mixes.ent`
- `NISO-supmatcustom-models.ent`
- `NISO-supmatcustom-modules.ent`

This follows the technique for making a new tag set suggested in the chapter [Implementing This Tag Set](#) in the JATS Journal Publishing Tag Library (follow the link [How To Make New Tag Sets](#) at the bottom of the page). Our custom files are included as [Supplementary Materials 1](#) and can be downloaded from <http://dtd.nlm.nih.gov/ncbi/supmatext/dtdFiles/>. We have also provided documentation in the form of a tag library can be found in [Supplementary Materials 2](#) as well as accessed through <http://dtd.nlm.nih.gov/ncbi/supmatext/extDocs>.

Elements that are not currently part of the JATS have been added in a namespace, “ext”. Although namespaces can be cumbersome we felt it was appropriate to use a namespace in this situation so that it would be easy to distinguish, and thus critique, concepts that are new with the extension.

Constraints for the Extension

In the process of creating the extension we developed a set of constraints to guide us in making modeling decisions.

First, because we were making an extension to JATS, we decided that information about supplementary materials should reside in the same xml document that describes the article of which they are a part. Although this might seem obvious within our context, we wanted to make this choice explicit because the [tagged samples](#) provided as Supporting Documentation to the NISO RP were delivered as stand-alone documents and it is possible that other applications might choose to keep this information in a separate file. This decision allowed us to drop the NISO required elements `<article-metadata>` and `<article-identifier>` because the information they encode is already covered by the elements found in the article’s `<front>` in JATS.

Additionally, we did not keep the NISO element <sup-mat-metadata>, which was used as the root element of the [non-normative DTD](#). There is not a direct analog to <sup-mat-metadata> in JATS, but it seemed to us that its main function, wrapping the supplementary data, could be mimicked by using <sec> to hold information about the supplementary materials, which is a common practice. Other information that was inside this wrapper, but outside particular pieces of supplementary material, could be covered by JATS elements. One of the few pieces of information inside <sup-mat-metadata> that we felt was not explicitly covered by JATS was the element <summary> (optional in the NISO RP). To cover this we added a new element <ext:summary> which we allowed inside <front>, if someone wants to write a summary of all the supplementary materials, and inside descriptions of individual supplementary material objects.

Second, because many of the elements and attributes found in the NISO non-normative DTD encoded information that was already found in JATS, we decided to favor JATS structures in these cases. For example, we retained the JATS models for <contrib-group> and <history> rather than replacing them with the non-normative DTD models for the same elements. Another example is the use of the JATS @mime-type attribute in preference to the non-normative DTD's <mime-type> element. Please see Appendix A for a more complete, but not exhaustive, list of these types of changes.

Finally, we decided that we would be willing to change names of elements or attributes found in the non-normative DTD (and listed in Table 1 of the NISO RP), and remodel some structures, if we felt that it would help us better understand and use the extension. While some examples of these changes are relatively minor, for example we changed the element name <provenance> to <original-sources> because we thought it better reflected the content of the element, some changes had broader implications and will be discussed below.

Rethinking Structure

The core of the NISO RP is the recursive element <sup-mat-object-metadata> which is used to represent information related to all levels of a piece of supplementary material. To do this the element can contain information that connects the supplementary material back to a specific spot in the article (in <article-component-metadata>), information that is descriptive (in <sup-mat-descriptive>), information about physical characteristics of the file (in <sup-mat-physical>), and information about subsequent levels (by allowing <sup-mat-object-metadata> to be recursive). The recursion of <sup-mat-object-metadata> is illustrated by a portion of one of the NISO tagged samples shown in Box 2. A schematic representation of the same code is found in Box 3, in order to condense the major concepts tagged using the NISO non-normative DTD.

Box 2A portion of a tagged sample from the Supporting Documents for the NISO RP showing the recursive nature of <sup-mat-object-metadata>.

The content of <article-metadata> has been removed to shorten the example. Also, some captions and titles have been shortened. The full version of this file is available at http://supplemental.niso.org/tagged-samples/10.1016_j.cub.2011.03.019-sup-mat-metadata-med.xml.

```

<!-- Supplemental materials contain both integral and additional objects -->
<sup-mat-metadata sup-mat-type="integral-and-additional" dtd-version="1.00">
  <article-metadata>redacted to preserve space</article-metadata>
  <sup-mat-descriptive>
    <summary>Supplemental Information includes Figures S1 and S2,
    Table S2, Supplemental Experimental Procedures, Acknowledgments, and
    Supplemental References embedded into a PDF file; Table S1 in the Microsoft
    Excel format; and Audio S1 and S2 in MP3 format.</summary>
  </sup-mat-descriptive>
<!-- PDF file with 6 embedded additional objects. These intellectually
distinct objects are all contained in one physical file. -->
  <sup-mat-object-metadata sup-mat-type="additional" relationship-of-
objects="unrelated">
    <sup-mat-descriptive>
      <summary>This PDF file contains six Additional objects: Figure
      S1, Figure S2, Table S2, Supplemental Experimental Procedures,
      Acknowledgments, and Supplemental References</summary>
    </sup-mat-descriptive>

<!-- NOTE that while physical metadata is provided for the 'parent' PDF, it
makes no sense to provide separate physical metadata for the embedded
'children' objects -->
  <sup-mat-physical>
    <file-metadata>
      <filename>PIIS0960982211002910.mmcl.pdf</filename>
      <format>PDF</format>
      <mime-type>application</mime-type>
      <mime-subtype>pdf</mime-subtype>
      <size>204 KB</size>
    </file-metadata>
    <ext-link href="http://download.cell.com/current-
biology/mmcs/journals/0960-9822/PIIS0960982211002910.mmcl.pdf"/>
  </sup-mat-physical>

<!-- Figure S1, related to Figure 1 -->
  <sup-mat-object-metadata relationship-of-objects="single">
    <article-component-metadata>
      <article-component-label>Figure 1</article-component-label>
      <article-component-title>Song Types</article-component-title>
    </article-component-metadata>
    <sup-mat-descriptive>
      <label>Figure S1</label>
      <titles>
        <title>Map of the Western...</title>
      </titles>
      <caption>Circles represent...</caption>
    </sup-mat-descriptive>
  </sup-mat-object-metadata>

<!-- Figure S2, related to both Figure 2 and Figure 3 -->
  <sup-mat-object-metadata relationship-of-objects="single">
    <article-component-metadata>
      <article-component-label>Figure 2</article-component-label>
      <article-component-title>Spectrographs of the Blue Song Type
      (Themes 23-30)</article-component-title>
    </article-component-metadata>
    <article-component-metadata>

```

```

        <article-component-label>Figure 3</article-component-label>
        <article-component-title>Spectrographs of the Dark Red
Song Type (Themes 31-37)</article-component-title>
    </article-component-metadata>
    <sup-mat-descriptive>
        <label>Figure S2</label>
        <titles>
            <title>Example Spectrographs of the
Evolution...</title>
        </titles>
        <caption>Theme 61 from the Light Blue...</caption>
    </sup-mat-descriptive>
</sup-mat-object-metadata>

<!-- Table S2, related to Figure 1 -->
    content removed for brevity

<!-- Supplemental Experimental Procedures -->
    <sup-mat-object-metadata relationship-of-objects="single">
        <sup-mat-descriptive>
            <titles>
                <title>Supplemental Experimental Procedures</title>
            </titles>
        </sup-mat-descriptive>
    </sup-mat-object-metadata>

<!-- Acknowledgments -->
    <sup-mat-object-metadata relationship-of-objects="single">
        <sup-mat-descriptive>
            <titles>
                <title>Acknowledgments</title>
            </titles>
        </sup-mat-descriptive>
    </sup-mat-object-metadata>

<!-- Supplemental References -->
    content removed for brevity

<!-- End of description for the objects embedded in the PDF file -->
</sup-mat-object-metadata>

<!-- Table S1, an integral object (an Excel worksheet) related to Figure 1 --
>
    <sup-mat-object-metadata sup-mat-type="integral" relationship-of-
objects="single">
        content removed for brevity

<!-- Audio S1, an integral object (MP3) related to Figure 2 -->
    <sup-mat-object-metadata sup-mat-type="integral" relationship-of-
objects="single">
        content removed for brevity

<!-- Audio S2, an integral object (MP3) related to Figure 3 -->
    <sup-mat-object-metadata sup-mat-type="integral" relationship-of-
objects="single">
        content removed for brevity

```


</sup-mat-metadata>

Box 3A schematic representation of the structure of the xml source from Box 2.

```
sup-mat-metadata (root)
  article-metadata
    sup-mat-descriptive
    sup-mat-object-metadata (PDF file)
      sup-mat-descriptive
        sup-mat-physical
        sup-mat-object-metadata (object in PDF file - figure)
          article-component-metadata
            sup-mat-descriptive
            sup-mat-object-metadata - (object in PDF file -
figure)
              article-component-metadata
              sup-mat-descriptive
            sup-mat-object-metadata - (object in PDF file -
table)
              article-component-metadata
              sup-mat-descriptive
            sup-mat-object-metadata - (object in PDF file - sup.
proc.)
              sup-mat-descriptive
            sup-mat-object-metadata - (object in PDF file -
acknowledgements)
              sup-mat-descriptive
            sup-mat-object-metadata - (object in PDF file - sup.
references)
              sup-mat-descriptive
            sup-mat-object-metadata (Excel file)
              article-component-metadata
              sup-mat-descriptive
              sup-mat-physical
            sup-mat-object-metadata (MP3 file)
              article-component-metadata
              sup-mat-descriptive
              sup-mat-physical
            sup-mat-object-metadata (MP3 file)
              article-component-metadata
              sup-mat-descriptive
              sup-mat-physical
```

Using what JATS has: <supplementary-material>

As we were trying to fit the NISO RP concepts into JATS we felt that the highest level at which <sup-mat-object-metadata> was used was mostly equivalent to JATS <supplementary-material>. As we looked further, we determined that the JATS <supplementary-material> model contained much, but not all, of the content found under <sup-mat-object-metadata>. We would need to add

content to the JATS <supplementary-material> model; some of this would be entirely new to JATS but other content existed within the JATS model, although it was not found inside <supplementary-material>. For elements that contained information new to JATS, such as <ext:original-sources> (<provenance> in the NISO RP) and <ext:preservation-level> (<preservation> in the NISO RP) we took their models directly from the non-normative DTD. For elements that existed in JATS, but weren't in <supplementary-materials>, such as <history>, we simply added the appropriate model to our extension.

We also realized that there were some wrapper elements under <sup-mat-object-metadata> that didn't seem necessary, or easy, to retain in our extension such as <article-component-metadata>, <sup-mat-descriptive>, and <sup-mat-physical>. This was partially because in some cases the content that would be wrapped inside them was already scattered throughout the JATS <supplementary-material> model. And since we had decided to change that model as little as possible they remained scattered and unwrapped. For example, we felt that the information found in <article-component-metadata> could be obtained through the use of @rid to reference other objects in the article, rather than needing to include a series of elements and repeating data found elsewhere.

Because <sup-mat-object-metadata> was recursive in order to describe the nesting of different physical or logical levels of a piece of supplementary material, there was going to need to be a way for the concept to also be recursive in JATS. Remember, we have already replaced the top-level <sup-mat-object-metadata> by <supplementary-material>, and it didn't seem at all right to let <supplementary-material> be recursive because that does not currently happen in JATS. This decision meant that we would need to consider a different hierarchy than what was used in the NISO non-normative DTD.

A new recursive structure: <ext:resource> and <ext:item>

As mentioned before, we used the NISO non-normative DTD extensively to try to understand the how the NISO RP could be used in practice. One surprising difficulty we had was that we would find ourselves getting lost in the recursive structure of the tagged samples even as we were looking at the publisher's version of the supplementary materials. This led us to believe that it might be helpful, for orientation, if there were distinctions to be made as you walked down the levels inside particular pieces of supplementary data. In particular, some levels had physical information associated with them (such as file size and file format) , while when you reached the level of describing a discrete piece of a single file (such as an individual figure in a pdf) this information type of information didn't exist. Yet these two types of object (physical and logical) were described by the same element <sup-mat-object-metadata>.

That there could be a tension created by using the same element to describe two different types of things is, likely unintentionally, pointed out by the comment just before the first occurrence of <sup-mat-object-metadata> in the xml sample in Box 2. It notes that in the instance of <sup-mat-object-metadata> following the comment it will contain physical metadata (describing the PDF, in <sup-mat-physical>), but that there is no reason to use <sup-mat-physical> for the subsequent child <sup-mat-object-metadata> elements because they are describing something that has no physical manifestation outside the enclosing PDF. We've tried to make this more obvious in our

schematic representation of the sample in Box 3 through the use of the label ‘object in’ inside the parenthesis.

We decided that it might be more intuitive to use two different elements; one to describe objects that have a physical presence (that can be addressed within a file system) and one to describe objects that can only be referred to (they are logical parts of an individual file). In the end we added two elements, `<ext:resource>` and `<ext:item>`, and deprecated `<sup-mat-object-metadata>`.

The generic element `<ext:resource>` takes the place of `<sup-mat-object-metadata>` below the level of `<supplementary-material>` and is allowed to be recursive. This element denotes the presence of, and wraps up information about, a physical object which is addressable in a file system. Examples of a resource are an individual file or a .zip file where the element would wrap the zip file, and then recursively wrap each file within that folder. Within the `<ext:resource>` wrapper is `<ext:resource-meta>` which holds all the metadata for the physical object, such as file name, size, and format.

The element `<ext:item>` takes the place of `<sup-mat-object-metadata>` when there is a need to describe a logical object inside a physical object. An example of this is the description of a PDF file, which contains three images, each with their own captions. The element `<ext:item>` would provide the opportunity for each of the image’s captions to be represented in the xml, at a level below the PDF file.

To walk through an example of a PDF file containing three images: the element `<supplementary-material>` is used to wrap the entire object; the element `<ext:resource>` is used to describe information about the overall PDF file; and the element `<ext:item>` contains information about each of the images in the PDF. Thus we’ve accomplished a recursing structure using our new elements, and provided some sign posts as to what level we may be at in the xml by using more specific names, like resource and item.

Because `<ext:resource>` is able to hold physical information about a file, there was no need for the JATS elements `<graphic>` and `<media>`, which would duplicate the same information, so they were removed from `<supplementary-material>` in the extension.

Box 4 provides an example of tagging using the extension. Notice how `<supplementary-material>` is the top level element, with `<ext:resource>` wrapping physically addressable objects and recursing when needed, and `<ext:item>` providing bottom level, logical information related to an object. A schematic version of the tagging is found in Box 5.

Box 4 Example of a piece of supplementary material tagged using the JATS extension.

This is the same information as is presented in Box 2. Some captions and titles have been shortened to keep the example concise. The full version of this file is available at [ENTER LINK HERE](#).

```

<!-- this article is at:
http://www.sciencedirect.com/science/article/pii/S0960982211002910 -->
<article xmlns:xlink="http://www.w3.org/1999/xlink" article-type="sample"
  xmlns:ext="http://jats.nlm.nih.gov/extensions/ext">
  <front>
    <journal-meta/>
    <article-meta/>
  </front>
  <body>
    <fig id="F1"/>
    <fig id="F2"/>
    <fig id="F3"/>
    <sec id="experimental-procedures">
      <title>Experimental Procedures</title>
    </sec>
    <sec>
      <title>Supplemental Information</title>
      <supplementary-material id="S1">
        <label>Document S1</label>
        <caption>
          <title>Two Figures, Table S2, Supplemental
Experimental Procedures, and Acknowledgments.</title>
        </caption>
        <ext:resource
xlink:href="http://download.cell.com/current-
biology/mmcs/....PIIS0960982211002910.mmcl.pdf"
          mimetype="application" mime-
subtype="pdf">
          <ext:resource-meta>
<ext:filename>PIIS0960982211002910.mmcl.pdf</ext:filename>
            <ext:format>PDF</ext:format>
            <size units="KB">204</size>
          </ext:resource-meta>
          <ext:item rid="F1" id="piece1">
            <label>Figure S1</label>
            <caption>
              <title>Map of the Western...</title>
              <p>Circles represent ...</p>
            </caption>
          </ext:item>
          <ext:item rid="F2 F3" id="piece2">
            <label>Figure S2</label>
            <caption>
              <title>Example Spectrographs of the
Evolution... </title>
              <p>Theme 61 from the Light Blue...</p>
            </caption>
          </ext:item>
          <ext:item rid="F1" id="piece3">
            <label>Table S2</label>
            <caption>
              <title>Sample Sizes for Number of
Singers...</title>
            </caption>
          </ext:item>

```

```

id="piece4">
    <ext:item rid="experimental-procedures"
        <caption>
            <title>Supplemental Experimental
Procedures</title>
        </caption>
    </ext:item>
    <ext:item rid="acknowledgments" id="piece5">
        <caption>
            <title>Acknowledgments</title>
        </caption>
    </ext:item>
</ext:resource>
</supplementary-material>

<supplementary-material id="S2">
    <!--description of Excel table, removed for brevity-->

<supplementary-material id="S3" rid="F2">
    <!--description of Audio 1, removed for brevity-->

    <supplementary-material id="S3" rid="F2">
        <!--description of Audio 2, removed for brevity-->

        </sec>
</body>
<back>
    <ack id="acknowledgments">
        <title>Acknowledgment</title>
    </ack>
</back>
</article>

```

Notice that the @rid attribute on <ext:item> is used to point back into the article, removing the need to use <article-component-metadata>. At this point we should also admit that we have basically ignored one of the main points of the NISO RP – the distinction between supplementary material that is integral versus that which is additional. We will discuss this further in the next section.

Box 5A schematic representation of xml source from Box 4 as a more visual representation of tagging with the JATS extension.

```

sec (title: Supplementary Material)
  supplementary-material (top level, logical information)
    label
    caption/title
    ext:resource (attributes: xlink:href, mimetype, mime-subtype)
(physical object)
  ext:resource-meta
    ext:filename
    ext:format
    size

```

```

        ext:item      (lowest level, logical information)
        label
        caption/(title, paragraph)
        paragraph
    ext:item      (lowest level, logical information)
        label
        caption/(title, paragraph)
        paragraph
    ext:item      (lowest level, logical information)
        label
        caption
        title
        paragraph
    ext:item      (lowest level, logical information)
        label
        caption
        title
        paragraph
    ext:item      (lowest level, logical information)
        label
        caption
        title
        paragraph
        supplementary-material (top level, logical
information)
        supplementary-material (top level, logical
information)
        supplementary-material (top level, logical
information)

```

Changes to JATS element: <fig>

The changes we decided to make to <fig> came from two different areas we had been thinking about at essentially the same time.

The first was that we liked the new (in relation to JATS) areas of content that the NISO working group had included in their description of supplementary material, such as explicitly labeling information related to preservation, separate authorship, or history of the object, and we felt that this information would be just as useful for describing non-supplementary materials, such as those that would be found in <fig>. Since we had the new model for supplementary material, and the models for <fig> and <supplementary-material> were similar in JATS, we opted to make <fig> in our extension match our model for <supplementary-material>.

The second was that we hadn't specifically addressed one of the required parts of the NISO RP; stating whether an object being described is considered integral or additional (or both) to an article. The NISO RP describes integral materials as those that are a necessary part of the intellectual work but that publishers, for technical, business or logistical reasons, have placed outside the core article. Additional materials are those that provide additional, relevant, or useful information related to the article, but are not necessary for understanding the core argument of the article. In the RP the integral/additional value is held on an attribute of <sup-mat-object-

metadata>. We made the assumption, for our extension, that what NISO RP refers to as additional materials could be marked-up with <supplementary-material>, which would leave elements like <fig> to be used to mark up integral materials. This fit in well with our decision to make the models for <supplementary-materials> and <fig> the same. Box 6 provides an example of tagging using the new model for the element <fig>.

Box 6 Example of tagging using the new <fig> model from the JATS extension.

```
<fig id="F1" xml:lang="en">
  <caption>
    <title>The fibrillin protein</title>
    <p> The fibrillin protein, mutated in Marfan syndrome, has about
60 domains, 47 of which...</p>
    <p>To see the <ext-link
xlink:href="http://www.ncbi.nlm.nih.gov/books/NBK22203....</p>
  </caption>
  <ext:summary>Contains two objects, a thumbnail image and higher
resolution image of the fibrillin protein mutation. </ext:summary>
  <alternatives>
    <ext:resource mimetype="image" mime-subtype="jpg">
      <ext:resource-meta>
        <ext:filename>fibrillin.jpg</ext:filename>
        <ext:format>JPG</ext:format>
      </ext:resource-meta>
    </ext:resource>
    <ext:resource
xlink:href="http://www.ncbi.nlm.nih.gov/books/NBK22203/bin/fib.prt">
      <ext:resource-meta>
        <ext:rendering-application>
          <ext:platform>
            <ext:software>
              <ext:software-name>CnD3</ext:software-
name>
            <ext-
link>http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml</ext-link>
          </ext:software>
        </ext:platform>
      </ext:rendering-application>
    </ext:resource-meta>
  </ext:resource>
</alternatives>
</fig>
```

Although we made this decision, and it is present in the DTD for our extension, we have since recognized that this assumption, that <supplementary-material> could be used exclusively for additional materials, and <fig> and other such elements could be used for integral materials, is weak. At the least this solution would have trouble marking up a single file that contained both additional and integral objects. And at the worst demanding that <supplementary-material> only contain additional materials might make use of the extension impossible for certain publishing

models. A possible way around this problem might be to use an attribute, such as @specific-use, on the elements <supplementary-material> and <fig> (as well as on <ext:resource>, and <ext:item>). We are uncertain if this solution would be adequate, or would merely add additional confusion rather than order.

Other Examples Marked Up In Our Extension

Although we realized that the extension was neither complete nor perfect, given the constraint that Karen's year as a Fellow was ending, we needed to stop working on the extension and so we brought our modeling exercise to an end. As a final exercise we retagged all the NISO Tagged Samples using our extension, as well as finding a few interesting examples 'in the wild' which would let us demonstrate using some of the optional elements that weren't used in the NISO Tagged Samples. To save space the tagged files can be found in SupMat 3 as well as at <http://dtd.nlm.nih.gov/ncbi/supmatext/sampleFileXML/>. We provide the following table to link the original articles to our files.

Table 2 Sample articles tagged using the JATS Extension.

Note that the entire content of the articles are not included, only those pieces which are needed to facilitate the tagging using the extension

Citation	Comments	File Name
NISO Examples		
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
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 <i>Oryza sativa</i> : 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
Huffard, C., R. L. Caldwell, and F. Boneka (2010), Male-male and male-female aggression may influence mating associations in wild octopuses (<i>Abdopus aculeatus</i>), J. Comp. Psychol. 124(1), 38–46, http://dx.doi.org/10.1037/a0017230		Octopus-ext.xml

Citation	Comments	File Name
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
Non-NISO Examples		
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://dx.doi.org/http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805500/		Adiposity-ext.xml
No current citation – forthcoming chapter on dbGaP from NCBI Handbook	Use of <original-sources>	Gap-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://dx.doi.org/10.3791/50191	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://dx.doi.org/10.1038/nature12131	Only described at level of PDF (caption offers good description)	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://dx.doi.org/10.1073/pnas.1211238109		Nitrous-ext.xml
Huang Liusheng et al. (2010). A modified method for determination of lumefantrine in human plasma by HPLC-UV and combination		Plasma-ext.xml

Citation	Comments	File Name
of protein precipitation and solid-phase extraction: application to a pharmacokinetic study. Anal Chem Insights. (5) pp. 15-23, http://dx.doi.org/10.4137/ACIS4431		
Cottrell, E., and Kelley, KA (2013) Redox Heterogeneity in Mid-Ocean Ridge Basalts as a Function of Mantle Source. Science 640 (6138), 1314-1317, http://dx.doi.org/10.1126/science.1233299	Use of <history>	Redox-ext.xml

Interlude

When we started this project we had assumed that we would create an extension of JATS incorporating the NISO RP for Supplementary Materials and simply offer it to the community as a starting point for discussion. But, as we tried to use the extension to tag sample files we found ourselves struggling to use it consistently, even though we had put a great deal of thought into renaming elements and into trying to be more specific about definitions of elements in the extension. Somehow a low-level friction, a tension, still existed. Though the analogy might be a bit strained, we felt that we had something akin to the creature from Mary Shelley's Frankenstein; while it had the ability to express complex concepts, maybe even quite well, it was outwardly ugly and its use felt clumsy at best. And while beauty and grace may not be typical qualities used to judge a document model, the utter lack of these qualities likely should be taken as a warning sign that all isn't quite right.

But, practically, where was the problem? In and of themselves neither JATS nor the NISO RP seemed to be the problem. JATS is used extensively and although one can certainly quibble with aspects of it it's not fundamentally broken or it couldn't be used the way it is. And though new, and although we did outline some changes we made to the non-normative DTD, nothing about the NISO RP seems fundamentally unsound either. The problem we were looking for seemed to be something deeper below the surface.

Shortly before Karen needed to present the results of her Associates project the phrase "supplementary material exceptionalism" coalesced in our minds as a way to describe this deeper problem. And this is what we will discuss in the remainder of the paper.

Part II – The Unintended Consequences of Creating the Extension

Supplementary Material Exceptionalism

By using the phrase "supplementary material exceptionalism" we mean that, currently, supplementary materials are primarily categorized by their supplementary-ness rather than by

what would otherwise be their dominant identifying characteristic, such as being a figure, a table, a movie, an external text document, etc.

During our work creating the extension, we were able to talk with some members of the committee which drafted the recommended practices, and we found that the idea of supplementary material exceptionalism is known to the community, though not everyone views it as an issue. There are those who feel that it is only the metadata regarding a supplementary object that needs to be present in the <supplementary-material> element, while there do seem to be others who say ‘I understand that it’s because of the way the element is defined, but, why are you stopping me from marking up my table or extended methods section in XML and placing it inside <supplementary-material>, inside my article XML document, where it would be cozy and safe?’ Everyone we spoke with, as well as ourselves, acknowledges that there is a great weight of practical business and production processes that will tend to slow any drastic change in direction. While acknowledging these challenges, we are also embracing our outsider status (neither of us can be said to work ‘in publishing’) to spur a conversation that promotes the sharing of all possible ideas whether brilliant or simply adequate.

We argue that it’s time to reexamine whether being labeled supplementary makes an object so exceptional that it needs to be treated wholly differently than it would be had it not been labeled supplementary. Or, another way of thinking about this might be that all objects should be treated with the care that the NISO committee gave to supplemental materials, as was alluded to in our discussion of making the model of <fig> match the model for <supplementalsupplementary-material> in our extension.

A Modest Proposal

What we are proposing is not a turn-key solution; it’s not a solution at all. Rather, it is a straw man that we hope will spur further discussion about this topic. (Even if that discussion mainly notes that the authors are a completely wrong-headed in their thinking and why, exactly, that is.) What this also means is that the following discussion is very general and very short. We highlight some of our ideas and provide a rough sketch of what a document using this scheme might look like, but we do not attempt to provide fine-grained model. In other words, there is no schema to be found in the supplementary materials section that represents these ideas.

1. Create a generic element that holds the information related to any block-level object that is allowed to float and does not directly reference an external file (and is generally referred to from the narrative text).

While one could move toward ending supplementary material exceptionalism by deprecating the <supplementary-material> tag and replacing it with other existing tags as appropriate, we don’t think that goes far enough. Rather, we suggest the use of a single generic element. For the rest of this paper we will represent this proposed generic with the tag <resource>, though we are not wedded to that name.

As the lengthy title of this section suggests, it is hard to succinctly define what objects would be covered by the generic element, though likely everyone has a gut-level feeling for what they

would be. If we use the definition given in the heading (block-level objects that are allowed to float but do not directly reference an external file) this would imply the deprecation of <boxed-text>, <chem-struct-wrap>, <fig>, <fig-group>, <preformat>, <supplementary-material>, <table-wrap> and <table-wrap-group> from JATS in preference to <resource>. Although they share the characteristics of being block-level and having the ability to float, the elements <graphic> and <media> reference external files which we believe puts them into a different category than the rest of the objects named in the last sentence (see 1 1/2 below).

As <resource> would be recursive, we don't think that there is the need for <resource-group> (in analogy to <fig-group> or <table-wrap-group>), but that is something that would need to be worked out in the course of practice.

One person we have spoken to about this is also interested in seeing appendices and lists included in the list of objects that might be deprecated. But we are somewhat uncomfortable with this, although we can't say exactly why.

Additionally, this generic element should contain, likely as optional elements, all the information (also mostly optional) that was defined by the NISO RP so as to capture as much information about the object as possible for current, and future, uses.

1½. Create a generic element that is used for referencing files outside the xml file

This is not absolutely necessary, but if there is a shift toward generics, then the elements <graphic> and <media>, having the same content model in JATS, could be replaced by a generic. Maybe <external-object>? It may be cumbersome, but it certainly generic and self-describing.

2. Create an attribute on the <resource> element that would signal its level of importance to the article

Okay – maybe it needs to be an element. But if <resource> is a generic, there needs to be a way to signal the level of importance of the object. This might be a binary distinction, such as integral v. additional from the NISO RP, or might evolve into a more nuanced series of values. We can also imagine that it could become a deeper set of classifications that might point to usage, like “raw data set”, “survey instrument”, “teaching module”, etc.

Conclusions

While the primary outcome of our project was a (usable) extension to JATS that supports the NISO RP, we ended up taking several theoretical detours and finally came up with a modest proposal to end supplementary material exceptionalism that we hope will spark discussion amongst the community. We fully recognize the need to remain within the bounds of reality by taking into account business and production practices, past attempts to work with supplementary material, and the varying theories and ideas of the community. We hope only to provide space for conversation and thought about what could be possible if we considered supplementary materials in the same way as objects within the narrative of the text.

Additional Resources

Supplementary Materials

- [SupMat1 – Custom DTD files for extension](#)
- [SupMat 2 – Tag library](#)
- [SupMat 3 – Extension sample files](#)

Web Resources

JATS - <http://jats.nlm.nih.gov/>

NISO RP - <http://www.niso.org/publications/rp/rp-15-2013>

NISO RP Supporting Documentation (includes Sample Tagged Examples and Tag Library) - <http://supplemental.niso.org/>

Acknowledgments

We would like to thank Jeff Beck, Alexander (Sasha) Schwarzman, Debbie Lapeyre and Kathy Kwan for interesting conversations related to the history and work of the NISO committee, as well as more targeted discussions both during and after our work on the extension.

This research was supported in part by an appointment to the NLM Associate Fellowship Program sponsored by the National Library of Medicine and administered by the Oak Ridge Institute for Science and Education as well as by the Intramural Research Program of the US National Institutes of Health, National Library of Medicine.

Appendix: Sample of Minor Changes Between the NISO Non-Normative DTD and the JATS Extension

This list should not be taken as exhaustive. We have tried our best to list the minor changes that would be of most interest to readers and which represent our general approach to the modeling.

- `<accessibility-long-desc>` is replaced by the JATS element `<long-desc>`
- `` and other face-markup elements are replaced by their JATS equivalents
- `<caption>` and `<titles>` as modeled in NISO is replaced by the JATS model for `<caption>`
- `<content-descriptor>` is replaced by JATS `@content-type`
- `<contrib-group>` as modeled in NISO is replaced by the JATS model for `<contrib-group>`
- `<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 replaced 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”.

This work is in the public domain and may be freely distributed and copied. However, it is requested that in any subsequent use of this work, the author be given appropriate acknowledgment.