June 2004

Proposed Tag Set Modifications for the NCBI Journal Archiving and Interchange DTD

Tag Set Modifications: Archiving and Interchange DTD

Here are some proposed alterations to the NLM Journal Archiving and Interchange DTD derived from suggestions sent via the suggestion web-form or submitted to the Archiving and Publishing discussion lists. This document will serve as the technical agenda for Tuesday's (June 8, 2004) conference call.

Each suggestion has been stated, accompanied by recommendation or discussion. In order to keep the conference call as brief as practical, suggestions have been divided into four categories:

- **1.0 Accept Without Comment** — Issue that (we hope) can be accepted as written. Any item can be moved to another category.
- **2.0 Discuss** — Issues that are more complex or more susceptible to multiple solutions where guidance from the Working Group is needed.
- **3.0 Philosophy Discussion** — Controversial items that may require more intense discussion.
- **4.0 Tag Set Documentation Issues**

1.0 Accept Without Comment

1.1 Add <initials> to <name> model

CHANGE: Add an optional <initials> element to the model of personal name <name>, changing the model from:

```
(surname, given-names?, prefix?, suffix?)
```

to

```
(surname, given-names?, initials?, prefix?, suffix?)
```

where initials is defined as:

```
<!ELEMENT  initials (#PCDATA %just-rendition;)* >
```

RATIONALE: Author's initials are useful in generating metadata for abstracting and indexing services that want initials instead of a given name.

CONSIDERATIONS: Documentation must make it clear that this element is NOT to be used to hold the middle initials of an author's full name. The element <given-name> should hold all "first names":

```
<name>
<surname>Smythe</surname>
<given-names>John Q</given-names>
<initials>JQS</initials>
</name>
```

1.2 Allow <named-content> anywhere face markup is allowed

CHANGE: Allow <named-content> anywhere that the text mark-up elements, like <italic>, <bold>, <ext-link>, <sup>, <sub>, etc are allowed, for example inside <source>, <preformat>,

and <keyword>.

RATIONALE: Several publishers currently use <named-content> to mark up, for example, genus-species names within text, and then make formatting decisions based on the presence of this tag (i.e. display italicized):

```
<nlm-citation citation-type="other">
<person-group person-group-type="author">
<name><surname>Yamazaki</surname>
<given-names>Y</given-names></person-group>
<source>Profiling of <named-content
content-type="genus-species">Escherichia coli
</named-content> chromosome (PEC)</source>
<year>2003</year>...
</nlm-citation>
```

1.3 Attribute changes for "`related-article-type`"

- **CHANGE**: Add "ext-link-type" to Related Article (with values like url, doi, etc.).

  RATIONALE: Make it possible to identify different types of related article links:

```
ext-link-type      CDATA        #IMPLIED
          Type of external link used to point to the
          related article. Use this attribute
          if the type of external link is one of the
          ones below or another named type.
          Suggested values include:
              ec          Enzyme nomenclature - see
                http://www.chem.qmw.ac.uk/iubmb/enzyme/
              gen         GenBank identifier
              genpept     Translated Protein-encoding
                          sequence Database
              highwire    HighWire press intrajournal
              medline     Medline or pubmed id
              pdb         Protein data bank. See
                          http://www.rcsb.org/pdb/
              pgr         Plant gene register. See
                          http://www.tarweed.com/pgr/
              pir         Protein Information Resource
                          see http://pir.georgetown.edu
              pirdb       Protein Information Resource
                          see http://pir.georgetown.edu
              pmc         Used to link between articles
                          in PubMedCentral access is
                          PMID
              sprot       Swiss-Prot. See
                        http://www.ebi.ac.uk/swissprot/
              aoi         Astronomical Object Identifier
              doi         Digital Object Identifier
              ftp         File transfer protocol
              uri         Website or web service
```

- **CHANGE/RATIONALE**: For internal PubMed Central processing, the Related Article element can already take the attributes "vol" and "page". Complete the potential information set by adding attributes for "issue", "journal-id", and "journal-id-type". For values for the journal ID type, use the same list as are already available

for the element Journal Identifier, plus “issn”.

```
The next five attributes are used to identify the
journal in which the related article was published.
vol         Volume of the journal in which the related
            article exists.  It is best practice to
            limit the scope to the current journal,
            the one in which the current article
            resides.
page        Page number of the related article. It is
            best practice to limit the scope to the
            current journal, the one in which the
            current article resides. The values for this
            attribute should be a first page or a page
            range.
issue       Issue number of the related article.
journal-id  Identifier for the journal which contains the
            related article.
journal-id-type
            Indicates whose identifier this is, for
            example, "pub-id" for a publisher's
            identifier or "pmc" for PUBMED Central.
```

- **CHANGE/RATIONALE**: Add attributes “id” and “alternate-version-of” to <related-article> to allow linking to multiple instances of an article.

  <related-article related-article-type=”commentary” ext-link-type=”pmid” xlink:href=”456789” id=”RA27”>Jones’s comments</related-article><related-article alternate-version-of=”RA27” ext-link-type=”doi” xlink:href=”/some/doi”/>

- Thus the full attribute set for Related Article would be:

```
"related-article-type CDATA  #REQUIRED
 ext-link-type    CDATA      #IMPLIED
 vol              CDATA      #IMPLIED
 page             CDATA      #IMPLIED
 issue            CDATA      #IMPLIED
 journal-id       CDATA      #IMPLIED
 journal-id-type  CDATA      #IMPLIED
 id               ID         #IMPLIED
 alternate-version-of
                  IDREF      #IMPLIED
 %might-link-atts;"
```

- **CHANGE:** Add the value “republished-article” as a suggested (potential) value for the attribute “related-article-type”. (Note: The attribute is defined as a CDATA attribute, so this is merely a documentation change.)

1.4 Add <issue-title> or <theme> to the article metadata

CHANGE: Change the article metadata model from:

```
article-id*, article-categories?,
title-group?, (contrib-group | aff)*,
...
contract-num*, contract-sponsor*,
conference*, counts?"
```

to the following, to add an element for the theme of an issue or a theme-related title for a special issue:

```
article-id*, article-categories?,
title-group?, (contrib-group | aff)*,
...
contract-num*, contract-sponsor*, theme?,
conference*, counts?"
```

where <theme> is defined as: (same content as a title)

```
(#PCDATA %struct-title-elements;)*
```

RATIONALE: Sometimes an issue is assigned a specific title, which represents the theme of the issue. This could be a special issue dedicated to a single subject, or an issue containing a collection of invited articles in a particular area.

1.5 Allow text in `<history>`

CHANGE: Change the model of <history> from:

```
(date+)
```

to the following:

```
(p | date)+
```

RATIONALE: When it is possible, it is preferable to use the <history> element with <date>s inside and all defined attributes. But sometimes sources blur all such distinctions and just present text. It would take a human being to interpret the text as the required values. It is better to preserve that text than not to preserve history at all or incorrectly.

```
<history>
<p>Paper submitted August 22, 2000; revised manuscript
accepted June  12, 2002.</p>
</history>
```

1.6 Attribute changes for "`pub-id-type`"

CHANGE: Add the values "pmcid" (pubmed central id) and "art-access-id" (generic article accession id for interchange and retrieval between archives) as suggested (potential) values for the attribute "`pub-id-type`".

1.7 Attribute changes for "`article-type`"

CHANGE: Add the values "bibliography", "opinion", "technical report", "dissertation", "communication" as values for the attribute "`article-type`". (Related discussion asks if this list should be open (#PCDATA) not enumerated. These would be added to the documentation whatever the decision.)

1.8 Add LaTeX

CHANGE: Add the ability to process LaTeX as well as TeX equations, and provide the Notation Declaration(s) necessary to make this change.:

1.9 Add <copyright-statement> to several models

CHANGE: Add a Copyright Statement <copyright-statement> to the models for:

- poetry <verse-group>
- extract <disp-quote>
- table <table-wrapper>

RATIONALE: This is done in books all the time, even if it is rare in journals..

1.10 Loosen `<citation> model`

CHANGE: Change the model for <citation>, which is currently:

```
(#PCDATA | %ext-links.class; |
 %label.class; | %references.class; |
 %simple-text;)*
```

to allow anything that may be inside a title to also be inside a citation. That includes:

```
%emphasis.class; | %inline-display.class; |
%inline-math; | %link.class; |
%math.class; | %phrase.class; |
%subsup.class;
```

RATIONALE: When a citation doesn't have the article title tagged, it needs to be able to handle anything loose within the citation that could be, in a more strictly tagged sample, enclosed inside the title tag. Currently, that is not possible.

1.11 `<inline-display>` versus `<inline-formula>`

CHANGE: Everywhere <inline-display> is allowed, allow <inline-math>.

RATIONALE: While the DTD was obviously designed to separate the two (a new class was created to make this possible), no rationale is given in the design notes or documentation, and we can't think of a good reason not allow both now either.

1.12 Widen models restricted now to face markup

CHANGE: Change the models that use %emphasis.class; (which contains just the face markup such as <bold> and <italic>) to include %emphasized-text; (which includes the emphasis elements, links, math, phrase elements, and subscript and superscript).

RATIONALE: The DTD (loose as it is) is actually too restrictive. In an archival DTD, what is the harm in loosening the models?

DISCUSSION: The current situation can lead to some very odd interactions. For example, the following <publisher-name> in a citation is illegal:

```
<publisher-name>
   Elsevier Science
   <ext-link ext-link-type="uri">www.elsevier.com</ext-link>
```

```
</publisher-name>
```

This is because <publisher-name> is defined as:

```
(#PCDATA "| %emphasis.class;  | %subsup.class;)*
```

However, the following is legal:

```
<publisher-name>
   Elsevier Science
   <bold>
   <ext-link ext-link-type="uri">www.elsevier.com</ext-link>
   </bold>
</publisher-name>
```

This is because:

```
<!ELEMENT  bold     (#PCDATA | %emphasized-text;)* >

<!ENTITY % emphasized-text
   "%appearance.class; | %emphasis.class; |
    %inline-math; | %link.class; |
    %math.class; | %phrase.class; |
    %subsup.class;">
```

2.0 Discuss

2.1 Non-Contiguous Page Ranges

CHANGE: The current DTD allows only for first page <fpage> and last page <lpage> when describing the metadata of an article (or reference to an article). This provides no way to say that an article started on page 14, continued until page 20, appeared briefly on page 34, and concluded on pages 38-40.

DISCUSSION QUESTIONS:

- Does this information need to be collected in an archive whose stated purpose is to preserve the intellectual content? This is information we may have. Is it information we need/want to preserve?
- If this information should be collected, should it be collected as positive information (this article runs on pages 14-20, 34, and 38-40) or as negative information: (This article runs from page 14 to page 40, skipping 21-33 and 35-37).

RECOMMENDATION: This information concerns presentation. The usual reason page numbers are collected is not to determine article length, but to provide an unambiguous identifier for the article when combined with such information as issue number and date.

Either:

- Do not capture such information in the XML; or
- If it must be captured, capture it as one string, in addition to the <fpage> and <lpage>.

RATIONALE: This has been seen in other journal DTDs.

CONSIDERATIONS: If captured as a string, it would be best to leave first and last page

alone and put the material in other element(s) since much citation matching depends on <fpage> and <lpage>.

Here is the way one DTD captured it, using negative ranges (excluded pages):

```
<!--       EXCLUDED PAGE(S)                        -->
<!--  Any page number(s) that must be excluded
      from the range defined by <fpage> and
      <lpage>.

      xpage requires two attributes:
        begin - the number of the first  excluded page
        end - the number of the last  excluded page

       (to exclude 1 page, @begin=@end)

      For the pagination 14-16, 19-22:
            <fpage>14</fpage>
            <xpage begin="17" end="18"/>
            <lpage>22</lpage>
      For 146, 148
            <fpage>146</fpage>
            <xpage begin="147" end="147"/>
            <lpage>148</lpage>

      For 100-101, 105, 107-120
            <fpage>101</fpage>
            <xpage begin="102" end="104"/>
            <xpage begin="106" end="106"/>
            <lpage>120</lpage>
                                                  -->
   <!ELEMENT xpage EMPTY>
   <!ATTLIST xpage
         begin   CDATA  #REQUIRED
         end     CDATA  #REQUIRED                 >
```

2.2 Changes To Model and/or Use Of <email>

Several requests came up in connection with the <email> element.

2.2.a CHANGE: <email> should be treated as just another kind of external link and allowed in all the places external links are allowed and given all the attributes and content of external links.

RATIONALE: Many DTDs handle email as an attribute value of External Link <ext-link>:

```
   <ext-link ext-link-type="email" xlink:href="nikos@atypon.com">
      <italic>nikos@atypon.com</italic>
   </ext-link>
```

It is in the spirit that an email is nothing but an external link anyway. The email address remains machine-processable (in an attribute), and the text retains original formatting (italics in this case in the content). This style also guarantees that an email address may appear anywhere a URL appears and vice versa.

2.2.b CHANGE: Allow <email> to be used inside the emphasis elements.

RATIONALE: In a source document that we wish to covert to Archiving, we have a chunk of text that has a small section that is italicized and has an email address embedded within it. Right now it is impossible to italicize the email address; it is necessary to "'bookend" the <italic> element around the <email> element and its content:

```
<author-notes>
<fn id="a1"><p><bold>Peter K. Piper</bold> <italic>Pickled
Pepper Farm, 1138 Middle St, Nowhere, Utah 12345
(<email>ppiper&#64;someemail.com</email>).</italic> Dr.
Piper is reknowned for his pickled pepper picking throughout
the...</fn>
</author-notes>
```

2.2.c CHANGE: Allow emphasis elements inside <email>.

RATIONALE: In a source document that we wish to covert to Archiving, we have several examples of email containing italics:

```
<footnote>
<p>Helpful comments from two anonymous referees on
earlier version of the paper are gratefully acknowledged.
Email: <externallink type="email"><i>kuse&commat;udel.edu</i>
</externallink></p></footnote>

<notes><p>Helpful comments from two anonymous referees on
earlier version of the paper are gratefully acknowledged.
Email:<email><italic>kuse&#x0040;udel.edu</italic></email>
</p></notes>2.
```

RECCOMENDATION: Don't add face markup to email, unless the change request **2.2.a** is also adopted. More to the point, preserving emphasis inside <email> is a look-and-feel issue. What content is added by knowing that the original journal placed email addresses in italics? What would it mean if only a portion of the address were italicized? 1) This DTD specifically makes it a goal to preserve the intellectual content and NOT the look and feel. That implies that some look and feel items may be lost. 2) The intention behind leaving email without face markup was to enable script processing to use the email to send a message. 3) In this case, the XML markup is entirely adequate to preserve the semantic intent and, depending on context, may be adequate to preserve the format as well.

2.3 Footnote Symbols

CHANGE: The current way to record a footnote symbol is using the "symbol" attribute of the <footnote>. Request that the footnote model be amended to allow a <label> element as well.

RATIONALE: While the symbol attribute can capture the actual symbol used, the <label> element is needed to capture "the exact display layout of the symbol used". The symbol attribute alone misses out on important rendering information. For example:

```
<fn id="fn1" symbol="*">
  <label><sup>*</sup></label>
  ....
</fn>
 <fn id="TF1-150"><label>12 </label>
```

```
<p>Data not available for 1 trial.</p>
</fn>
```

DISCUSSION: 1) This DTD does not attempt to capture "the exact display layout". 2) An unstated design constraint for the DTD Suite is that there should *not* be two ways to do something unless there are good reasons for the difference. There is a way to tag this data now. To introduce another would permit two mechanisms. To introduce another and get rid of the first is not a backward compatible change.

2.4 Custom Metadata

CHANGE: Add a new element Custom Metadata <custom-meta> to the model for both journal and article metadata to allow the preservation of information not specifically named in the Archiving DTD.

```
<!-- ============================================================= -->
<!--                    CUSTOMIZED METADATA ELEMENTS               -->
<!-- ============================================================= -->

<!--                    CUSTOM  METADATA                           -->
<!--                    User-defined, auxiliary metadata data field.
                        Used to capture metadata elements that
                        have not been defined in the models for this
                        DTD, so that no intellectual content will be
                        lost. It will most probably be used for
                        special cases, product-specific, or other
                        non-standard metadata, for example the
                        journal-history information preserved in at
                        least one publisher's DTD.                 -->
<!ELEMENT  custom-meta  (meta-name, meta-value)                     >

<!--         used-by    Explanation why/when/how/by whom this metadata
                        has been used
             xlink:href There are many linking attributes, which
                        will rarely be used.  The most important is
                        "xlink:href", which holds a URI such as a
                        URL or filename to link to a term online.  -->
<!ATTLIST  custom-meta
             used-by    CDATA                                #IMPLIED
             %might-link-atts;                                       >


<!--                    NON-STANDATRD METADATA, DATA FIELD NAME    -->
<!ELEMENT  meta-name    (#PCDATA | xxx)*                  >


<!--                    NON-STANDATRD METADATA, DATA FIELD VALUE   -->
<!ELEMENT  meta-value   (#PCDATA | xxx)*                  >
```

RATIONALE: Some publishers include considerably more metadata with their SGML/XML articles than the current DTD allows. We suggest a flexible mechanism for allowing additional metadata about both the journal and the article.

DISCUSSION: What should be inside the new elements, as a replacement for the "xxx" above? There are several possibilities, including:

- At least %link.class; and %phrase.class, or alternatively,

- All of:

```
<!ENTITY % emphasized-text
    "%appearance.class; | %emphasis.class;  |
     %inline-math; | %link.class; |
     %math.class; | %phrase.class; |   %subsup.class;" >
```

This introduces the rather curious idea of face markup in metadata. For most metadata, this seems like a bad idea. But for some metadata, for example, an author's bio, which is metadata but also gets printed and displayed, face markup is probably appropriate.

DISCUSSION: Suggested placement for this new element could include:

- As the last element in the journal metadata <journal-meta> (optional, repeatable),
- As the last element in the article metadata article-meta <article-meta> (optional, repeatable),
- As the last element in Front Matter <front> (optional, repeatable).

DISCUSSION: Should this new metadata element repeat as needed in the places named above, or is there a need for a grouping tag to hold the set?

```
<!ELEMENT  custom-meta-wrap  (custom-meta+)   >
```

2.5 Another way to indicate role in citations

CHANGE: Add <role> to the content of <person-group>:

RATIONALE: In a source document that we wish to covert to Archiving, we have discovered several examples of sample data that include role information in citations.

```
Source tagging:
  <citation id="i0012-9615-073-02-0259-johnson1" HyTime="bibloc">
  <author itype="norm">
  <surname>Johnson</surname>
  <fname>D. H.</fname>
  </author>
  <author itype="norm">
  <fname>T. A.</fname>
  <surname>O'Neil</surname>
  <role>managing directors.</role>
  </author>
  <date>2001</date>
  <othinfo>Wildlife habitat relationships in Oregon and
  Washington. Oregon State University Press, Corvallis,
  Oregon, USA</othinfo>
  </citation>

Suggested Archival Tagging:
    <citation id="i0012-9615-073-02-0259-johnson1">
    <person-group>
    <name>
    <surname>Johnson</surname>
    <given-names>D. H.</given-names>
    </name>
    <name>
    <surname>O'Neil</surname>
```

```
<given-names>T. A.</given-names>
</name>
<role>managing directors.</role></person-group>
<year>2001</year>
<named-content content-type="othinfo">Wildlife habitat
relationships in Oregon and Washington. Oregon State
University Press, Corvallis, Oregon, USA</named-content>
</citation>
```

DISCUSSION: There is already a way to tag this information, using the "person-type-group" attribute on <person-group>:

```
<person-group person-group-type="managing directors">
```

- Is there material in <role> for an author that cannot be handled with this attribute?
- Do we need both methods?
- If both are needed, how will a conversion house distinguish between them?

RECOMMENDATION: Do not add this element to the references. An unstated design constraint for the DTD Suite is that there should *not* be two ways to do something unless there are good reasons for the difference. There is a way to tag this data now. To introduce another would permit two mechanisms. To introduce another and get rid of the first is not a backward-compatible change.

2.6 Attribute values for "`article-type`"

CHANGE: Change the attribute values for "`article-type`" from an enumerated list to an open CDATA.:

RATIONALE: In an Archival DTD, why not?

3.0 Philosophy Discuss

3.1 Untitled Sections

CHANGE: The DTD Suite defines a section as a headed unit of text. There have been several requests that titles be made optional.

RATIONALE: The publishing DTDs from which people are converting allow untitled sections, so there is source data with untitled sections.

DTD BACKGROUND: The Archiving DTD is built with the definition of section as headed text. This was done deliberately to facilitate web display and automatic table of contents generation, which we viewed as two of the major functions for this DTD.

The decision at design time was that this would negate the ability to retain unheaded sections, but that:

- Untitled sections were rare;
- Untitled sections are often confusing to the reader; and
- Blank titles and untitled sections do not work well in online generated navigation tools.

In our collective experience, untitled sections have been used mostly as a typographic convention and to control page layout. If the sections are also logical divisions, they would most probably have titles.

DISCUSSION of SAMPLES: Prior to this meeting, a call went out for examples of untitled sections. The request was that both visual (html, PDF, fax) and code be provided. The samples received fell into three categories:

- Case 1: Unheaded initial section
- Case 2: Unheaded sections that contain nothing but other sections

and we could envision one other possibility, often seen in magazines, although no samples were provided:

- Case 3: Unheaded sections in Editorial/Commentary-style content

Case 1: Unheaded Initial Section ? **Problem**: An unheaded section before the start of the first real (headed) section, usually used as an introduction. **Analysis:** No problem. This is not a section, this is just paragraph-level material before the first section in the body. **Solution:** No DTD change required. To tag this material, delete the <section> wrapper, leave its content loose at the beginning of the <body>.

Case 2: Unheaded Sections Containing Nothing But Other Sections ? **Problem**: An unheaded section enclosing nothing but another section or sections. **Analysis:** We examined the formatted pages for these to see how was this handled. What did this look like in print? In print, it was not detectable, all the sections had titles. How could this be? The empty sections were being used to force the section head into being a heading level 3 instead of the larger heading level 2. This is not really an unheaded section, this was a way of faking out the print engine to get the formatting someone wanted. **Solution:** There is a way to tag such sections in the current DTD. The outer-level section is removed. The inner-level section is given a display-level ("disp-level") attribute on <section> that can be set to "3" or "third" or "h3" or whatever the software needs to make it work.

Case 3. Unheaded Sections In Editorial/Commentary-Style Content ? **Problem:** These are common enough in magazines that no one can swear that they do not exist in journals, although no one sent any examples. These really are untitled sections. In print, they show up as some extra whitespace, which serves to break up the print page nicely. On the website version of the same story, they do not show up at all. **Solution(s):** There is no way in the current tagset to preserve these, but we would like to argue that it does not matter. What intellectual content has been lost if the three white-space-delimited sections become 7 paragraphs instead?

DISCUSSION: Has anyone else seen any other cases of untitled sections? Are the cases we have seen sufficient to make section titles optional? What are the benefits of doing so? What are the penalties?

(Note: Since titles are not optional, people admit that they have been committing tag abuse an inserting empty title elements. Whether or not we make section titles optional, do we need to suggest a solution for this problem?)

3.2 Preservation of Look and Feel

BACKGROUND: There have been several changes requested that are really questions about the nature of the archival record being preserved. Answering them (positively or

negatively) makes a statement about the philosophy and basic nature of the XML archive for which this DTD Suite is intended.

There are (at least) two basic ways an XML archive can seek to preserve its content.

- One way eliminates (as much as possible) specific look-and-feel decisions of original text (for one example, the punctuation between keywords), marking up the semantic content (<keyword>), and letting a stylesheet generate the formatting (such as the commas or semi-colons that separate the keywords). Archives may behave this way 1) in the interest of ensuring that the articles in the archive present as uniform a face as possible to the user (and that distinctions of format are meaningful, not just branding), and 2) to keep the archive as semantically clean as possible for searching.
- A second archival strategy implicitly states that you must make no assumptions concerning rendering or stylesheets for the XML and that all look and feel decisions should be preserved in the XML. This means that there must be a way to preserve whether the keyword list used semicolon, colons, or periods between the keywords. For example, an <x> element could be added to the DTD to hold such punctuation in a keyword list:

```
<keywordgroup>
<heading id="h1001" level="5" format="inline">Key
words<x>.</x></heading>
<keyword>Agricultural landscape</keyword><x>, </x>
<keyword>Akaike weights</keyword><x>, </x>
<keyword>amphibians</keyword><x>, </x>
<keyword>geographical range boundary</keyword><x>, </x>
<keyword>habitat loss and fragmentation</keyword><x>, </x>
<keyword>mammals</keyword><x>, </x>
<keyword>niche breadth</keyword><x>, </x>
<keyword>vertebrates</keyword><x>.</x>
</keywordgroup>
```

THE REQUESTED CHANGE: By design, the current NLM DTD provides limited support for encoding punctuation and formatting information in the XML. Some DTDs make it a point to include special optional tags with the sole purpose of preserving punctuation, boilerplate text, and other *important rendering information* directly in the XML.

There are a number of places in a DTD where punctuation and formatting information could be preserved, but where it is not allowed by the current models:

- authors,
- citation references,
- keywords,
- definition lists.

RATIONALE: In the requestor's opinion, preserving such formatting is considered preferable to relying on such information being furnished by a stylesheet. Documents become standalone and as a result document management and rendering processes are simplified . Especially for large archives, content from several publishers with different conventions and rendering styles can be treated in an homogeneous manner and displayed with a common stylesheet. The presence of an explicit <x> tag allows for easy removal of formatting and punctuation information, when these are not needed. This could be the case when collecting metadata information for an article. <x> tags are supposed to contain

punctuation, spaces and other non-content bearing information, whose removal does not affect the intellectual content of the document.

RECOMMENDATION: This DTD Suite was written according to a specifically stated goal *not* to preserve the look and feel of an original article, that the important aspect was to preserve the intellectual content. We suggest not adding <x> tags.

4.0 Tag Set Documentation

We have not listed the bug fixes here; those changes will just be made when new documentation is generated as a result of the versioning of the DTD.

4.1 Accept Without Comment

4.1.1 Attributes: Display "Values" before "Remarks"

CHANGE: In the Attribute Section, the Remarks (which detail how and why an attribute can be used and often contains lists of "suggested" attribute values) is currently displayed on the page *before* the actual attribute value are listed. The suggestion is to reverse the order, so that the values will be discussed first. Additional language that makes explicit when attributes are from a restricted list and are not CDATA would also be helpful.

RATIONALE: The documentation of attributes confusing. When reading quickly or trying to look something up, it is not always obvious whether an attribute is CDATA or from a list. (For an example, in the attribute "pub-type", the remarks section lists many "suggested" values, but the text describing the value says that values are completely open (not limited).

4.1.2 Modify Tag Library navigation bar design

CHANGE: Change the design of the left-hand navigation bar to make it easier to identify each of the items in it. Perhaps alternating light and dark backgrounds behind the items, or alternating text colors.

RATIONALE: Since some of their names are several words, and therefore wrap, it can be easy for people to miss items in the list. Experienced users are not finding items in the middle of the Tag Library, and, in some cases, are suggesting we add sections to the Tag Library that are already there.

4.2 Discussion

4.2.1 Print version of the documentation

DISCUSSION: The last Working Group meeting did not rate a print version of the Tag Library as a very high priority. We decided that most people will look online to take advantage of the linking. Requests have been made since. Has the priority changed?

4.2.2 Element Context Placement

CHANGE: Replicate the relevant portion of the Context table that shows the contexts of an element (where an element can be used) on each element description page.

RATIONALE: Both what is inside an element and where you can use an element will be in one convenient location.

DISCUSSION: There are several questions:

- Is this a good idea?
- Would a link to the appropriate spot in the current context table be sufficient? Better or worse?
- If this information is put on the element pages, should the current Context Table be kept or deleted since it is now duplicate information?
- How do these decisions effect a potential print version? (Context tables are LARGE.) Even if we want the information in both places for the electronic documentation, we may not want it twice in print.