Tag Set Working Group Minutes
June 2007

Table of Contents

1. Tag Set Working Group Meeting

The Working Group meeting took place by conference call on June 14, 2007. This meeting was to set the guiding principles to be used in evaluating non-document-compatible change requests. Before the next meeting a list of potential changes, with suggested resolutions based on these principles, will be presented to the Working Group.

The next meeting will be in middle to late August, 2007. Please send your available dates to Jeff Beck so that he can schedule a time convenient for most.

1.1 Attendees

1.1.1. NLM Staff
- Jeff Beck (Moderator)
- Abraham Becker
- Laura Kelly
- Moira Lee
- Adeline Manohar
- Kim Tryka

1.1.2. Secretariat (Mulberry Technologies, Inc.)
- Deborah A. Lapeyre

1.1.3. Telephone Participants
- Alex Brown (Griffin Brown)
- Beth Friedman (DCL)
- Linda Good (Cadmus)
- Evan Owens (Portico)
- John Meyer (Portico)
- Bruce Rosenblum (Inera)
- Parke Shissler (Cadmus)
- B. Tommie Usdin (Mulberry)

2. Recommended Rules of Thumb for Making Changes

Although version 3.0 tag sets will be not be document-backward-compatible, the changes will be guided by the following.

- Structures added in version 3.0 should be optional not required.

- Requested changes that will make the models better should be approved, for example, adding additional wrappers to group existing elements or new functionality. The rule of thumb is "If the change is good for the tag sets, let's do it!"

- If a change would require data that many people do not have, we will not make it or will make the new content optional.

- Check carefully for potential granularity mismatches before making a change, and change granularity only if necessary.

3. Usage Guideline Documentation

Scope: A P A B

Recommendation: Tag Library documentation should add discussions (possibly in essay form) for many elements and element combinations (such as <contrib> and <aff> or <fig>, <fig-group>, and <graphic>) that describe effective practices for solving common problems. Where the industry has several common ways to tag material, they should be illustrated, explaining the differences between the styles and providing tagged examples of each. These are not "best practice" recommendations but "common encoding" examples. By providing acceptable alternative practices, we hope to head off some of the tag abuse seen in the wild. It is also important by both implication and example to illustrate how NOT to tag structures where common errors are known and can be described.

Suggested topics include:

- Citations
- <contrib> and <aff>
- <fig>, <fig-group>, and <graphic>
- <table>, <table-wrap>, and <table-group>
- Book and product reviews
- Letters to the editor with replies
- Short news subarticles inside a main article

4. Recommendations

4.1 <pub-date>

Scope: A P B

Recommendation: <pub-date> should remain a privileged data and not part of <history>. Samples of multiple publication dates should be added to the documentation.

4.2 <x>

Scope: A

Recommendation: Allow <x> "everywhere you can." This will include inside <contrib> and <contrib-group>.

4.3 <citation>

Scope: A P A B

Recommendation: Make three citation models:

- The current <citation>, which is a #PCDATA model with any of the reference elements in any order as many times as needed with nothing enforced;
- The current <nlm-citation> (thoroughly documented that this is for a specific client and is not necessarily widely applicable);

- A new all element citation, that allows all the elements in the other citation models in any order and repeatable. In Green, this will also include <x>.

4.4 Layout

Scope: **A** **P** **A** **B**

Discussion: We needs some formatting and layout control in these tag sets ─ but not too much. Defining "too much" eludes us. Like pornography, we know it when we see it, but were not sure that can we could pin it down with a definition. The Working Group reconfirmed that the tags sets are not designed to support production directly, users who want to do this will require customizations. The rules of thumb decided by the Working Group are 1) that if the group-of-practice represented by the Working Group needs it, then we add it and 2) the tag sets add only what "we have to," trying to stay on the structural and semantic side of the line and avoid the purely presentational.

Specific Recommendations:

- Since this one keeps being requested, add an orientation attribute to table (and the rest of the display objects such as figure).

- Add a phrase-level element similar to <named-content> whose purpose is to record typographic style variation. (Do not make this a milestone element, non-overlapping stylistic hierarchies will have to repeat elements inside the formal hierarchy.)

4.5 Interleaving <p> and <sec>

Scope: **A** **P** **A** **B**

Discussion: The Working Group notes that there are both book and article models in the wild that permit loose paragraphs after sections have begun. The Working Group decided that allowing title-less sections was sufficient support for this questionable structure.

5. Detailed Discussion

5.1 Multiple versions of a single object

Scope: **A** **P** **A** **B**

Request: There is a need for a housekeeping structure to hold multiple forms or versions of a single object. The type case is a graphic for which multiple resolutions or multiple formats exist (such as a tif, a gif, a jpeg, and a png). The user is not aware of these groupings; they are for processing and software. Other potential use cases include:

- An XML-tagged table, the same table as a csv file, and a spreadsheet of the same table for user experimentation;

- Two versions of a section, one for print and a significantly longer one for online;

- A video and a thumbnail that represents one image in the video;

- An equation present as MathML, TeX, and a jpeg;
- Two captions for a figure, one for color and one for black and white.

Recommendation: Investigate the possibility of one or more wrapper elements to encase multiple versions of an object. There are a number of objects which exhibit this property. Note: This is a different issue from that of multiple objects placed together in a group (such as a figure group) so that they can be cited together (see next item).

5.2 Multiple objects linked to together in a group

Scope:

Request: The journal literature is full of multiple objects grouped into a unit, for example groups of figures and groups of tables and other clustering that are not explicit groups, such as a group of individually numbered, captioned graphics lumped together in a single figure, groups of equations cited with a single equation number. These are groupings that are directly experienced by the user.

Recommendation: Review the grouping and clustering mechanisms in the tag sets with a view to the possibility of:

- Regularizing the names of groups and wrappers;
- Adding additional clustering mechanisms;
- Providing samples of grouped objects to illustrate several ways it can legitimately be accomplished (many graphics in one figure, many figures in one figure-group, etc.).

5.3 <contrib-group>

Scope:

Request: For Green, should <contrib> be completely unordered, allowing preservation of anyone's presentation order? If the model is loosened, documentation should be provided to illustrate the most common ways publishers are using it.

6. Add to the Discussion List

6.1 Supplementary Material

Scope:

Request: The spreadsheets, datasets, audio files, and other material that is more and more frequently associated with the "text" of an article. Publisher's practice, even the definition of supplementary material, varies widely. Our working definition is "the stuff that publishers want to accompany an article that does not print." Supplemental material is not inside a paragraph and the reference to it might not be inside one either, it may be, for example, inside a table.

The Working Group separated supplementary material from the problem of multiple forms of material (the jpeg, the gif, and the png for a graphic). That may still leave several classes of such material:

- Material that is not part of the narrative textual flow of an article, but that was used to support the conclusions of the narrative, e.g., a dataset and a survey for a paper that presents the highlights learned by the survey and recorded in the dataset.

- Material added for enhancement purposes, such as a video of a reaction that is also described with text and still images.

- Material that is logically part of the published material but that is not printed or displayed, e.g., 10 columns of a 45-column table are displayed. Is this really just an alternate form, as a video might be an alternate form of a still image?

6.2 <person-group>

Scope: **A** **P**

Request: Even in the loose <citation> model it is not possible to reorder the person's name or to punctuate it, except inside <string-name>. Some publishers would like to reorder the name to match their presentation order.

6.3 Magazine elements

Scope: **A** **P** *A*

Request: Many journals are beginning to be published including more magazine-style content. Some publishers have modified the NLM tag sets for magazines. Add magazine metadata and structures to the NLM tag sets.

Recommendation: Provide a list of potential elements and content model changes so that the group can determine if this is just adding elements to the tag sets or if this is a new variant "Pink" tag set.

6.4 Issues of journals

Scope: **A** **P**

Discussion: By design, the NLM tag sets represent journal articles and do not represent journals. There are objects in journals that are not associated with any article, such as the Table of Contents, the cover picture, the list of the advisory board, display advertising, etc. Do we need to model this material? The Working Group can revisit this issue after version 3.0 is cooked.

6.5 Manifest or packaging

Scope: **A** **P** *A* **B**

Recommendation: The Working Group can revisit this issue after version 3.0 is cooked.

-------------------- document end ----------------------