NISO STS (Standards Tag Suite) Technical Working Group Minutes for STS Draft Version 0.9

For NISO STS Technical Working Group
February 2017

February 01, 2017, 10:00 am - 11:00 pm EDT

1 Introduction

These are the minutes for the Technical Working Group meeting for the NISO activity to create a standard tag set for Standards. Details on this work item are available to the public at:

`http://www.niso.org/workrooms/sts/`

These minutes record decisions made and action items assigned during the NISO STS Technical Working Group call on February 01, 2017.

1.1 Attendees for February 01, 2017

- Dreyfuss, Bob (ASTM: observer)
- Flanagan, Heather (RFC)
- Galichet, Laurent (ISO)
- Gooskens, Frans (NEN)
- Gupta, Vinay (Edaptive Technologies)
- Hollowell, Bob (ASME)
- Imsieke, Gerrit (le-tex Publishing Services)
- Juillerat, Serge (ISO)
- Lapeyre, Debbie (Mulberry Technologies)
- McRae, Mary (IQ Solutions)
- Markantonatos, Nikos (Atypon)
- Hollowell, Bob (ASME)
- Rosenblum, Bruce (Co-chair, Inera)
- Saari, Antti (SFS: invited expert)
- Usdin, Tommie (Mulberry Technologies)
- Wheeler, Robert (Co-chair, ASME)
- Winchell, David (XSB)

1.2 Administrative Business

- Minutes from the previous meeting were accepted as posted.

- Bruce Rosenblum reminded us the next stage is a second draft of the Tag Set and documentation for public voting.

1.3 NISO STS Technical Working Group Call

The NISO STS Technical Working Group will meet by conference call on Wednesday February 08, 2017, at 10:00 am US EST.

US and Canada Toll free number and Passcodes are available at

`http://www.niso.org/apps/org/workgroup/sts-technical/`
under "Upcoming events".

Global numbers can be found via http://bit.ly/1KCHbsT
For security reasons, the Passcode will be required to join the conference.

2 New and Continuing Action Items

For All

- **@xml:lang** — The Working Group has recommended that @xml:lang be allowed on the elements <kwd-group> and <subj-group>, to name the language in which the child elements are written. The xml:lang attribute has *not been proposed* for individual keywords or subjects within these groups (following JATS practice). Members who publish in multiple languages are requested to check their multi-language documents to determine if all of the subjects or keywords in a group are in a single language, or if keyword or subject groups can contain more than one language.

- **@vocab-version** — This attribute was requested because some members are using versioned taxonomies that change rapidly. Mulberry recommends that we follow the JATS Standing Committee decision and *not add* a @vocab-version. Several members (Ivan, Laurent, et al.) take an action item to answer the questions: Do the taxonomies you are using in your standards documents provide a version identifier in either the name or the URI? (GMDM, ITF 7, etc.)

For Serge Juillerat, Laurent Galichet, and Ken Rawson (in abstentia, sorry Ken)

- **ICS Codes** — Carefully read section 3.9.5 and consider: Do you wish to preserve the <ics> element long term (we will preserve <ics> in the short term for backward compatibility)? Are you comfortable with the bifurcated solution in which some ICS codes are in <ics> and some are in <subject>? What are the ramifications of approaches #1 or #2 (bifurcation) or #4 (new grouping element) for your data going forward (future migration, transforming <ics> into <subj-group>, etc.)?

For Gerrit Imsieke

- **@xml:lang** — The Working Group has recommended that @xml:lang attribute be allowed on the elements <kwd-group> and <subj-group>, to name the language in which the child elements are written. The @xml:lang attribute has *not been proposed* for individual keywords or subjects within these groups (following JATS practice). Gerrit Imsieke took an action item to determine if all of the subjects or keywords in a group are in a single language, or if groups can contain more than one language.

- **IPC Codes** — The Working Group recommended that the IPC codes be tagged as subjects. Gerrit Imsieke will report on the following questions:

  - Should the IPC subject codes be tagged as @vocab="ipc" or as @subj-group-type="ipc"? That is, is "ipc" a @vocab name that should take a @vocab-identifier?

  - Should any of the vocab attributes be used for IPC codes, even if we decide that "ipc" is a subject group type?

- **UNSPSC Codes** — The Working Group recommended that the UNSPSC codes and UNS codes be tagged as subjects. Gerrit Imsieke will report on the following questions:

  - Should UNSPSC subjects be tagged as @vocab="unspsc" or as @subj-group-type="unspsc"? Is "unspsc" a @vocab name that should take a @vocab-identifier?

- Should any of the vocab attributes be used for UNSPSC codes, even if we decide that UNSPSC is a subject group type?

For Debbie Lapeyre

- **Vocab attributes** — Send Robert Wheeler some examples of the use of the @vocab attributes in JATS.

For Robert Wheeler

- **@subj-group-type** — Find or create an example of a subj-group/@subj-group-type.

- **Vocab attributes** — Find or create an example of a kwd/@vocab and a @kwd/@vocab-term-identifier. Find or create an example of a subject/@vocab and a @subject/@vocab-term-identifier.

- **Meta Note Attributes** — Assess if the following attributes are sufficient for the new element <meta-note>. (These are the current attributes for <normative-note>, <non-normative-note>, and <notes-group>.)
    - @content-type (Type of Content)
    - @specific-use (Special purpose or usage)
    - @id (Document Internal Identifier)
    - @originator (Originating standards organization)
    - @xml:base (Base)

- **Request #00703** states that ASME includes a number of blurbs on their copyright pages. These are currently tagged as <sec>s inside <front> and further defined with @sec-type. ASME requested some additional suggested @sec-type values to cover these blurbs. Now that there is a new <meta-note> element, should these blurbs be tagged as <sec>s or as <meta-note>s? Robert Wheeler will assess whether Request #00703 is resolved by adding <meta-note>.

- **Request #00707** — For a number of American SDOs there is a standards "authority" (that may be comparable to ISO for NSBs?). It seems we do not have a place (in metadata) to represent this material, e.g., a text string that may be on the cover and title page such as, "An American National Standard". When time ran out there was agreement that this text is not <secretariat>, but no agreement as how it should be tagged. Robert will develop a straw-man recommendation to answer the following questions: Is this information a <meta-note> with a type attribute of "authority" (or similar value)? Is this a separate piece of metadata, not a note at all? Is this text related to ISO's <release-version> or not?

- **<release-version>** — The <iso-meta>, <reg-meta>, and <nat-meta> elements may contain a <revision-version> element to indicate values like "IS" (International Standard), "DIS" (Draft International Standard), etc. Robert Wheeler will revisit the original SDO metadata discussions and report the original thinking on whether this element or a similar element is needed for the metadata elements <std-meta> and <std-doc-meta>.

3 Comments Requiring Technical Recommendations

Many of the Comments on the NISO STS draft, as submitted through the NISO website comment form, were trivial changes or typos and could be fixed without discussion. Some Comments could

be resolved by the editor (Mulberry) or by the editor in consultation with the original requestor. Other Comments require discussion by the NISO STS Technical Working Group plus invited experts. Those Comments requiring technical discussion are described in this section.

In the February 1st 2017 meeting, only item *3.9 Keywords and Subjects* was discussed.

3.1 Resolved: ASTM-15: <elocation-id>

See earlier minutes.

3.2 Resolved: ASTM-26: Expanding where <proj-id> may be used

See earlier minutes.

3.3 Resolved: #00667: Footnote type values

See earlier minutes.

3.4 Resolved: #00669/#00685: Citation to a Standard (<std>) and its model

See earlier minutes.

3.5 Resolved: #00668/#00685/ASTM-19: <glyph-data>/<private-char>

See earlier minutes.

3.6 Resolved: From #00685: Amend model of <std-ref>

See earlier minutes.

3.7 Resolved: #00682: <adoption> with no interior document

See earlier minutes.

3.8 Resolved: #00689: Rationalize Table of Contents models

See earlier minutes.

3.9 Tabled: #00690: Keywords and Subjects

3.9.1 Resolved: #00690-a: Keywords and subjects defined

Recommendation

- NISO STS will take both the keywords elements (container <kwd-group>) and the subject elements (container <subj-group>) from JATS. The content models will not be changed.

- Keywords will be defined as specific terms or concepts found within the text of a standards document, or implied (broader, narrower, etc.) from terms found within this text. Even though keywords may be found in the text of the standards document, the source of the keywords could be a taxonomy, a thesaurus (https://www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en:term:2.62), a controlled vocabulary (https://www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en:term:2.12), or an uncontrolled vocabulary.

Although similar in content to index terms, tagged keywords are part of the metadata (inside a *-meta element), not scattered throughout the prose of the standards documents.

- Subjects will be defined as overarching categories, classifications, topics, or themes of a standards document, as a higher-level descriptor of the content. In JATS, subjects are used to organize articles in a Table of Contents or similar. In standards documents, subjects have been used to organize standards into series. The source of the subjects can be a classification scheme (https://www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en:term:2.6), such as a taxonomy.

3.9.2 Resolved: #00690-b: Keywords and subjects attributes

- The elements <kwd-group> and <subj-group> will take the following attributes:
    - @id (Document Internal Identifier)
    - @specific-use (Specific Use)
    - @xml:base (Base)
    - @vocab (Name of a controlled or generic vocabulary, taxonomy, database, etc. that is the source of the term. e.g. "CRediT", "inspec", or "structural engineering")
    - @vocab-identifier (unique ID and possibly pointer to the vocabulary, typically a URI or DOI reference to this vocabulary)
    - Type attributes;
      - The JATS attribute @kwd-group-type will be retained, with the JATS definition changed to omit "the name of an ontology, thesaurus, index or other term source where the keywords in this group are defined", which will now be tagged using @vocab. Organizations such as "IEEE" and "ASME" will be tagged as @originator. The usage of kwd-group-type will reflect an authority that assigned the keywords to the standards document, such as, for example, "committee" or "working group". Typically, in cases such as these where there is no named vocabulary, the @vocab attribute should also be set to "uncontrolled".
      - The JATS attribute @subject-group-type will be retained for <subj-group>
    - @originator (Which standards producing organization placed these keywords or subjects *into the document*, particularly useful in jointly authored standards. The @originator attribute does *not* name the originator of the term or the owner of the vocabulary, that is either @vocab for controlled standards or @kwd-group-type for uncontrolled ones.)

- The children of elements <kwd-group> and <subj-group> (<kwd>, <nested-kwd>, etc.) will take the following attributes:
    - @id (Document Internal Identifier)
    - @specific-use (Specific Use)
    - @xml:base (Base)
    - @vocab-term (Canonical term as expressed in the named vocabulary. The content of the <kwd> or similar element might not be exactly the same as the canonical form; the <kwd> content might be a user-specific variant.)
    - @vocab-term-identifier (unique ID and possibly pointer to the specific term in the named vocabulary, typically a URI or DOI reference, but this could be an item number or other identifier.)

3.9.3 Resolved: #00690-c: Keywords and subjects Best Practice examples

- The INSPEC keyword examples from IEEE and ASME will be added to the documentation.
- The following example of a controlled vocabulary will be used in the documentation.
```
<kwd-group vocab="VWL">
```

3.9.4 Tabled: #00690-d: Additional keywords and subjects attributes

3.9.4.1 Tabled: @xml:lang

Recommendation

The Working Group has recommended that @xml:lang be allowed on the elements <kwd-group> and <subj-group>, to name the language in which the child elements are written.

Action

- The xml:lang attribute has *not been proposed* for individual keywords or subjects within these groups (following JATS practice).

- Gerrit Imsieke took an action item to determine if all of the subjects or keywords in a group are in a single language, or if groups can contain more than one language. Members who publish in multiple languages are requested to check their multi-language documents as well.

Mulberry Recommendation

- Allow @xml:lang only at the group level, not on individual subjects and keywords.

3.9.4.2 Tabled: @vocab-version

This attribute was requested because some members are using versioned taxonomies that change rapidly.

Discussion

The versioning of taxonomies, ontologies, and vocabularies is a very complicated thing that is currently in flux. It is not obvious if the vocabulary version needs to be a number, a date, a time-stamp, child elements, multiple attributes, or what. The Data Citation community may have ideas and should be consulted.

Additionally, in the current climate, more and more vocabularies are putting their version information into their names or URIs, so the version would be captured in the @vocab or the @vocab-identifier attributes. For example, Creative Common licensing, you can point to a specific version of the license by URI. Similarly, different version of the Dublin Core have different URIs, and the base URI is used to establish @vocab-term-identifier as well.

Mulberry Recommendation

- Follow the JATS Standing Committee decision and *not add* a @vocab-version attribute.

Action

- Several members (Ivan, Laurent, et al.) take an action item to answer the questions:
  Do the taxonomies you are using in your standards documents provide a version identifier in either the name or the URI? (GMDM, ITF 7, etc.)

3.9.4.3 Tabled: IPC codes

Recommendation

IPC codes will be tagged as subjects, and the following example used in the documentation:

```
<subj-group>
  <compound-subj>
     <compound-subj-part content-type="code">B82B1/00</compound-subj-part>
     <compound-subj-part content-type="value">Nano structures
   </compound-subj-part>
 </compound-subj>

 <compound-subj>
   <compound-subj-part content-type="code">H01L21/02</compound-subj-part>
   <compound-subj-part content-type="value">Manufacture or treatment of
   semiconductor devices or of parts thereof</compound-subj-part>
 </compound-subj>
</subj-group>
```

Action

However there remain a few questions:

- Should the example above be tagged as @vocab="ipc" or as @subj-group-type="ipc"? That is, is "ipc" a @vocab name that should take a @vocab-identifier?

- Should any of the vocab attributes be used in this example, even if we decide that IPC is a subject group type?

- Gerrit Imsieke takes these questions as an Action Item.

3.9.4.4 Tabled: UNSPSC codes

Recommendation

The working Group recommended that the UNSPSC codes and UNS codes be tagged as subject codes.

Action

- Should these be tagged as @vocab="unspsc" or as @subj-group-type="unspsc"? That is, is "unspsc" a @vocab name that should take a @vocab-identifier?

- Should any of the vocab attributes be used in for UNSPSC codes, even if we decide that UNSPSC is a subject group type?

- Gerrit Imsieke takes these questions as an Action Item.

3.9.5 Unfinished: #00690-e: ICS Codes

When time ran out, the Working Group was working on how ICS Codes should be encoded.

Recommendation

- For reasons of backwards compatibility, we cannot remove the current <ics> #PCDATA-model element.

- For reasons of backwards compatibility, we cannot rename <ics> either.

Discussion

There are several alternatives for how we proceed from here:

1. **Bifurcation** — We could create a bifurcated approach, where <ics> would be used for non-hierarchical ICS codes (numbers only) and <subj-group> would be used if the user wanted to include the ICS descriptions (titles) or to build a hierarchy of ICS codes.

   This approach has the substantial disadvantage that there would be two *very* different elements to tag the same ICS codes material. Documentation would need to make both ways very clear and indicate cross references between them:

   ```
   <ics>42.42.42</ics>
   ```

   versus

   ```
   <subj-group subject-group-type="ics">  (or maybe @vocab="ics")
    <compound-subj>
      <compound-subj-part content-type="code">42.42.42</compound-subj-part>
      <compound-subj-part content-type="value">Poodles</compound-subj-part>
    </compound-subj>
   </subj-group>
   ```

2. **Temporary Bifurcation** — We could create a bifurcated approach, exactly as in option #1, but for Best Practice would recommend the use of subject groups for ICS codes. The <ics> element would be first dis-recommended, then deprecated, and finally phased out, and only subject groups would remain.

3. **Codes include text** — We could leave ICS Codes as text only as they are currently modeled (#PCDATA) and place both a code and its description (title) as text inside <ics> as well: (*Mulberry note: This is a bad idea and no one spoke in its defense during the call.*)

   ```
   <ics>13.110 Pressure Seals, Gasogenes, and Poodles</ics>
   ```

4. **#PCDATA Plus** — We could remodel <ics> (currently #PCDATA) as #PCDATA plus inclusions for code and description elements, so the same element could be used in three ways: either as #PCDATA, as element content, or as mixed-content. (*Mulberry note: This is a very bad idea, but it was mentioned briefly during the meeting.*)

5. **Grouping** — We could leave the current <ics> as it is (with a model of #PCDATA and a definition that says ICS Code-only) and create a new grouping element, specific to ICS codes, such as <ics-group>, with a model such as:
   ```
         ics-group  (ics, ics-desc?)
   ```
   Then, wherever <ics> is now, we would allow the grouping
   ```
        (ics  |  ics-group)*
   ```
   The OR choice above would be allowed inside all the *-meta elements. If <ics-group> (with a better name, please) needs to be recursive, the model could be expanded (similar to subject groups):
   ```
         ics-group  ((ics, ics-desc?), ics-group)
   ```
   or
   ```
         ics-group  ((ics, ics-desc?), ics-group*)
   ```

Possibly pertinent points made during the call:

- For some of you, ICS codes, although metadata, get printed out in the content, as many of you print lists of keywords.

- For one of you: "In our own metadata, we do not use the supplied ICS codes, we go one level deeper when we enrich the XML file. We would need the subject group construct, and need it to be recursive."

- In the original STD Metadata Recommendation Report, the Subcommittee did not intend for the element <ics> to be included in the <std-meta> or <std-doc-meta> elements (as it is in the current draft), the Subcommittee wanted ICS codes to be tagged as subject groups.

- For ISO at least (and possibly for others), there is some discomfort at the thought of moving away from a dedicated ICS element, but you say you might be able to get used to it. [IEEE, which also has a dedicated element, was not on the call.]

- One of you expressed the opinion that ICS codes are critical, but ICS titles (descriptions) should not be preserved in the XML as this is static content and typically a lookup.

Mulberry Recommendation

Mulberry recommends the ICS grouping model (whatever we call the elements). This solution has several advantages:

- No change need be made to current ISO STS and ISO-related standards documents, they could use <ics> as they do now.

- Each NISO STS user could choose whether to use just the ICS code values or to include the (admittedly typically static) ICS code description (title). Many current users prefer to include both the code and the description (title), since ICS codes change over time.

- ICS codes would always found in the same place in the metadata, and *they would use the same tag*.

- Several of you think that ICS codes are privileged information that is important in the standards world, not just an ordinary classification scheme, and this solution would preserve that distinction at very little disadvantage to those who think ICS codes are just subject classifications.

- The new <ics-group> would need to be inside all the *-meta elements, including <iso-meta>, <reg-meta>, and <nat-meta>. With <ics> in all the metadata groups, ISO and ISO-related standards organizations could switch to <ics-group> at any time or keep doing as they did under ISO STS.

- For IEEE, who currently uses both a code and a description, a simple name transformation would change their current elements
        ```
        ics  (ics-code, ics-desc?)
        ```
  into
        ```
        ics-group  (ics, ics-desc?)
        ```

- A bifurcated solution might be hard to explain and justify during public review. And bifurcation means that users will have to know to look in two places for ICS codes.

3.9.6 Roads Not Taken

- The @content-type attribute will not be added to <kwd-group> or <subj-group>, since the required functionality can be obtained using @kwd-group-type or @subj-group-type and the new vocabulary/taxonomy attributes.

- The @originator attribute will not be placed on the children of <kwd-group> or <subject-group>.

- NISO STS will follow the JATS example of using @content-type for individual terms and subjects but the more specifically-named @kwd-group-type and @subj-group-type for the container elements. There is a virtue in consistency.

- Other classification schemes that can found in standards document (IPC, DK, ICS, UDC, etc.) do not need to be shown in examples in the documentation.

- It was suggested and rejected that such ICS descriptions (titles) might be preserved in attributes, since attributes could not be used to make a hierarchy.