NISO STS (Standards Tag Suite) Differences Between ISO STS 1.1 and NISO STS 1.0

Version 1

October 2017

1   Introduction

The NISO Standards Tag Suite (NISO STS) 1.0 is based on ISO STS 1.1 and is fully backward-compatible with it. In extending it to meet the needs of a variety of other Standards Development Organizations (SDOs), the NISO STS committee added significantly to the already large tag set. This document describes the types of additions made and some of the reasons for the additional structures and loosened content models.

1.1 Four NISO STS Tag Sets

NISO STS includes four implementations of the suite, called "Tag Sets"

- the Interchange Tag Sets (one Tag Set with MathML 2.0 and one with MathML 3.0), and

- the Extended Tag Sets (one Tag Set with MathML 2.0 and one with MathML 3.0).

These tag sets are built from the elements and attributes defined in the Suite and are intended to provide models for standards publishing and interoperability. The Extended Tag Sets differ from the Interchange Tag Sets in allowing the use of OASIS Exchange (CALS) tables in addition to XHTML tables.

1.2 Relationship of NISO STS to ISO STS

NISO STS is based on the International Organization for Standardization (ISO) Standard Tag Set (ISO STS, version 1.1). Although the superstructure of ISO STS was left intact, additional elements and attributes were added to create NISO STS; both new standard-specific elements and attributes and newer JATS elements and attributes. In addition, much optional material was added to existing ISO structures.

NISO STS is a fully backwards-compatible superset of ISO STS 1.1. No ISO STS elements/attributes were deleted in making NISO STS 1.0, although some required elements/attributes have been made optional. Any document that is valid to ISO STS 1.1 will be valid to NISO STS 1.0, by changing only the document type declaration (or other schema identification mechanism) to point to NISO STS instead and by changing the DTD version attribute on the top level element (<standard>). (Note: This assumes conversion of ISO STS to one of the DTDs, XSDs, or RNGs that use the 2.0 version of MathML. ISO STS uses MathML 2.0, and MathML 2.0 and MathML 3.0 are not backward compatible.)

1.3 Relationship of NISO STS and ISO STS to JATS

ISO STS 1.1 was based on the Journal Article Tag Set (JATS 1.0: ANSI/NISO Z39-96-2012), therefore NISO STS is based on JATS as well. What NISO STS and ISO STS took from JATS is their basic textual structures. Thus block structures (such as paragraphs, figures, tables, and equations), bibliographic references, floating structures (such as equations, statements, and footnotes), inline face markup (such as bold, italic, and underline) are all the-same-as or completely-compatible-with JATS markup. Standards metadata is entirely different, so ISO STS has its own standard-specific metadata and NISO STS built in that metadata, adding even more metadata tags than ISO STS already had.

NISO STS 1.0 has been brought up to JATS level 1.1 (ANSI/NISO Z39-96-2015) with selected material from JATS 1.2d1 included as well. (JATS 1.2d1 was not completed in time for full inclusion in NISO STS 1.0.)

1.4 Extent of Changes from ISO STS to NISO STS

This report describes the changes made to ISO STS to create the new NISO STS. Both the standards-metadata and the underlying JATS element/attributes changed extensively between ISO STS 1.1 and NISO STS 1.0. *Every element* in ISO STS has changed its content model, added attributes, or both. If no other change was made, JATS 1.1 and above assign an optional @xml:base attribute (new to ISO STS) and optional @id attribute to every element. No elements have been deleted, an entire infrastructure has been added, and new metadata elements have been added.

As an analogy, consider the framework for a house. A foundation is made, then a wooden framing structure is built. Then walls, floors, a roof, windows, lights, furniture, and landscaping are all added. The original frame holds it all together, but the frame can hardly be seen for the layers on top. So the original ISO STS may be hard to recognize, under the layers of new material, except by XML parsers, which know that all valid ISO DTD documents are still valid to NISO STS 1.0.

These changes are so extensive, that making an element-by-element comparison between the two tag sets would be long, involved, difficult to read, and ultimately rather like a Schema difference report produced by software, too minutely comprehensive to be useful. Thus this report describes the nature and extent of these changes, but does not report on a content-model-by-content-model basis.

1.5 Non-normative Documentation

The non-normative Tag Libraries for NISO STS 1.0 are entirely new. Even the material, such as selected element descriptions that was taken from the ISO STS documentation (with kind permission from ISO) has been extensively rewritten. Nearly all of the examples are new.

2 Changes/Enhancements

2.1 Adoptions

While many standards documents are newly created by a standards-producing agency, many other standards are adopted from a different standards-producing agency. When a standard is adopted, the adopting organization may wrap their own metadata, front matter, and/or back matter around the original standard, producing a combined, nested document. As an example, the standard "DIN EN ISO 1012" contains an ISO standard (ISO 1012), which was adopted by EN (EN ISO 1012), and then the adoption was adopted by DIN (DIN EN ISO 1012). The outer wrapper adoption contains information created by DIN (including superseding notes), which is wrapped around the adoption created by EN, which in turn is wrapped around the ISO standard.

One of the goals of NISO STS was to allow (as much as possible) the textual components of a standards document to be tagged in reading order, that is, in the order in which the user sees them. A new element Adoption (<adoption>) was created as a document-level element that may contain its own metadata, front matter, and back matter, and that will contain, in addition, one or more complete standards or other adoptions.

An <adoption> may contain:

- Its own front matter including metadata, back matter, and then

    - the full text of the standard being adopted,

    - the full text of another adoption ,

    - a reference to a standard that is elsewhere (This is particularly useful when a standards organization does not possess the XML version of the standard they are adopting, but only, for example, the PDF version or in situations where the adopting organization does not have permission to distribute the standard being adopted.), or

    - or an <xi:include> element, to bring in a standard or adoption being managed in a separate file.

- Following the content of an adoption, it is possible to adds adopting back matter element. provided by the adopting organization, to contain annexes in the local language, additional requirements, or notes.

Some adopting organizations place all their adoption material, including adoption-specific back matter, *before* the standard they are adopting; others place the adoption-specific back matter *after* the standard being adopted. The NISO STS model for adoption allows either location to be used, though not both at once.

2.2 New Standards Metadata

One of the goals of the NISO STS effort was to produce a metadata model for standards documents that would meet the needs of most SDOs. Design goals for this new metadata model included the following.

- The model should be enabling not enforcing. This means that organizations desiring tighter control will need to create a more restrictive subset or restrict content in another processing layer, for example, using Schematron.

- The new metadata model must not interfere with current ISO STS processing; any document valid to ISO STS 1.1 should still be valid to current NISO STS.

A new metadata holder NISO STS Standard Metadata (<std-meta>) was created to hold metadata specific to or produced by a standards-producing organization (such as ISO or an SDO) that is publishing or adopting a standard. The new metadata element contains all of the non-ISO-specific elements named within the three ISO-related metadata elements, new standards-related elements (many of which were also added to the ISO-related three, and new elements from JATS.

Each <std-meta> element contains the metadata for *one* standards-producing organization and may be used to describe:

- organizations such as SDOs that are not part of the ISO infrastructure and thus have never used any of the three ISO-related metadata elements, or

- any ISO-related standards organization, for which <std-meta> can replace any of the three ISO-related STS metadata container elements.

2.3 Enhanced Standards Metadata

The three ISO STS metadata elements (ISO Metadata <iso-meta>, Regional Metadata <reg-meta>, and National Metadata <nat-meta>) have been enhanced with the descriptive power that was created with the NISO STS standard metadata element. Most of the new material provides additional description of the standard.

Descriptive material from JATS has been added to <iso-meta>, <reg-meta>, and <nat-meta>:

- abstracts,
- keywords,
- subject categories,
- self URI, and
- the rest of the JATS counting elements (table count, figure count, word count, etc.).

New metadata that was defined for NISO STS standard metadata (std-meta>) has also been added to <iso-meta>, <reg-meta>, and <nat-meta>:

- a descriptive name and location information for the publisher/adopter of a standards document (in the same way that the element <publisher> in JATS provides name and location information concerning the publisher of a journal article) (A further grouping element allows for there to be more than one publisher/adopter),

- identification of the Technical Committee, Subcommittee, and optionally Working Group responsible for the standard (For example, "ISO/TC 27/SC 4". These may be nested for hierarchical committee descriptions.),

- notes that are part of the metadata not part of the text of the standard, and

- the more complex title parts added for NISO STS standard metadata.

2.4 Avoiding Recording Information Redundantly

Particularly in an adoption, the same piece of metadata information can be added at several levels inside a standards document, for example, if the abstract(s) for a standard are the same for all levels of the metadata (inside both the ISO and the Regional standards body metadata, for example), then the abstract would be the same inside both metadata containers. To reduce this duplication of metadata information within a standards document, a new overarching metadata element (Standard Document Metadata <std-doc-meta>) was created to hold metadata that pertains to the entire standards document. Thus, in the case just cited, the abstracts can be placed inside the element <std-doc-meta> and thus not placed within the ISO or the Regional standards body metadata, but treated *as if they existed* for both ISO and Regional versions of the standard.

By design, the elements inside <std-doc-meta> are elements that standards organizations may share. Any element within <std-doc-meta> should be counted by an application as being part of any other standards metadata-wrapper within the document. Any element within <std-doc-meta> may be overridden by placing the same element within any of the four organization-specific metadata elements: <std-meta>, <iso-meta>, <reg-meta>, or <nat-meta>. Thus, to continue the example, if a National standards body has its own unique abstract, that will be placed inside the National standards body metadata (<nat-meta>).

The new document-level metadata element is optional, so an organization can choose to ignore this layer completely.

2.5 Non-TBX Terms and Definitions

NISO STS provides two very different structures for tagging the terms and definitions inside a Term Section (<term-sec>). One way is to tag terms using TBX (a namespaced vocabulary based on ISO 30042; TermBase eXchange). The second way is to use <term-display>, a simpler but less powerful alternative which provides a looser description of a term and its definition than TBX-tagging allows, using natural language (potentially containing just text) to describe terms.

NISO STS loosened and simplified the content of the <term-display> element (making a large undifferentiated OR group) and added additional semantic tags, to make tagging existing terms and definitions easier than using the more strictly structured TBX tagging, while being backward compatible with ISOSTS V1.1. The new <term-display> model:

- enables tagging term and definition content *in the sequence in which that content appears* in the standards document,

- enables, but does not require, tagging of the principle semantic components of terms (such as the term itself, its definition, related-terms, part of speech, pronunciation, term source, etc.), and

- enables tagging that will make it possible to extract terms and definitions for use in glossaries containing the terms and definitions from many standards.

2.6 Editing ACE (Amendments/Corrigenda/Errata) Documents

Many standards bodies produce discrete, standalone amendment/corregenda/errata (ACE) documents. Corrigenda are complete documents that may modify either a standard or an amendment. For some standards-defining organizations Amendments and Corrigenda are the same; for others they are not.

An ACE document needs the same metadata as a standard, plus metadata that describes the document being amended. In addition to the small, nearly trivial documents (e.g., a paragraph of instructions followed by a paragraph to be inserted), an Amendment can be quite large, e.g., a new annex or a complicated multi-level-nested section.

For creating standalone ACE documents, a new element was introduced to hold the instructions to a reader on how this amended text relates to the existing text of the standards document being amended or corrected. The element Editing Instruction <editing-instruction> holds the text that describes how an alteration or addition should be made to the standards document. These instructions are typically placed before the portion of the document to be modified. For example, an editing instruction might contain the words: "insert this new paragraph as the first paragraph of section 2.3.1" or "delete the paragraph below from section 2.3.1".

2.7 Book Structures

2.7.1 Structural Indexes

Two types of indexing elements are included in the NISO STS Tag Set:

- elements for describing a structural, multi-level index (typically used for capturing a back-of-the-standard or similar index that is already present in a standard being converted to this

Tag Set, although it may also be used for indexes generated from index terms set within the text of the standards document), and

- elements for tagging inline index terms within the text, to be used to construct an index.

2.7.2 Structural Tables of Contents

Elements for describing a structural, multi-level Table of Contents were included in the NISO STS Tag Set. Such tagging is typically used for capturing a Table of Contents already present in a standards document being converted to this Tag Set, although it may also be used for Tables of Contents generated from standard and adoption metadata.

2.7.3 XInclude

The XInclude mechanism is provided in NISO STS, so that standard and adoption documents can be managed as separate files and "included" (<xi:include>) as needed into a final document.

2.8 Additions from Recent JATS

2.8.1 MathML 3.0

ISO STS includes only MathML 2.0. Both the Interchange and Extended versions of NISO STS are available using either MathML 2.0 or MathML 3.0.

2.8.2 Additional Elements

Elements were added to NISO STS to contain the following.

- Programming code and schema fragments, including attributes to describe the language, language version, type of code, identify executable code, etc.

- Additional granularity in addresses (city, state, postal code)

- Structures for citing datasets, e.g., the formal title or name of a cited data source (or a component of a cited data source) such as a dataset or spreadsheet. A data title element (for the name of the data piece being referenced) was added to the list of citation-content elements, for inclusion in citations, product descriptions, et al. A version element was also added to the list of citation-content elements to hold "the precise version number of the data or software being cited".

- An externally defined identifier for a person such as a contributor or principal investigator. This element could hold, for example, an ORCID, a trusted publisher's identifier, a JST (Japan Science and Technology Agency) identifier, or an NII (National Individual Identifier).

- An externally defined institutional identifier, whether publisher-specific (for example "AIP") or from an established identifying authority (for example, "Ringgold" or "ISNI"). A single institution may have multiple identifiers, either because different assigning agencies have defined an identifier or because the entity belongs to more than one larger organizational unit.

- Licensing and open access information (from the NISO ALI recommended practice)

- Alternatives for multiple expressions of citations and collaborations

- More content allowed inside table cells, for example paragraphs.

2.8.3 New Attributes from JATS

Attributes from JATS that were added to NISO STS include the following.

- The attribute @xml:base was added to *each element* in the Tag Set for use in providing a base URI for identifiers in an XML-tagged standards document or a part of an XML-tagged standards document.

- The attribute @id (a type ID attribute) was added to any element that did not already have an ID-type attribute, so that all elements may be uniquely identified within their document.

- **Vocabulary/Taxonomy attributes** — Added four new attributes for naming (and possibly linking to) a general or controlled taxonomy, ontology, vocabulary, index, database or other source of terms. These attributes may be used on keywords, subjects, terms, etc.

  - vocab — Name of the controlled or uncontrolled vocabulary, taxonomy, ontology, index, database, or similar that is the source of the term. For example, for a subject term, a value might be IPC Code ("ipc") or UNS Code ("uns"). For an uncontrolled term. the value might be an area of study such as "structural engineering" or merely the word "uncontrolled".

  - vocab-identifier — Unique identifier of the vocabulary, such as (but not limited to) a URI or DOI.

  - vocab-term — The content of the element which takes this attribute is the free prose version of the vocabulary or taxonomic term. The @vocab-term attribute holds the canonical version of the same term, as it appears in the vocabulary.

  - vocab-term-identifier — Unique identifier of the term within a specific vocabulary, such as (but not limited to) an item number, a URI, DOI, etc.

2.9 Added Structures

2.9.1 Standard-specific Structures

Standard-specific elements (not based on JATS or BITS) added to NISO STS include the following.

- **Notes** — All notes were non-normative in ISO STS; NISO STS added normatives notes. Notes may now be grouped under a heading in a Note Group structure.

- **Examples** — All examples were non-normative in ISO STS; NISO STS added normatives examples.

- **Standards Titles** — Because the titles of standards can be complex, the three possible components of a standards title (introductory, main, complementary) have been expanded. Each of the three title components can now contain a label, a title of the correct sort, and a subtitle, allowing for many more combinations. (Note: the full title resides in an element of its own.)

- **Standard Identifying Information** — There are many additional NISO STS elements and attributes for providing standard identification metadata. Multiple designators (i.e., dated and undated) are allowed for a standard. ISBNs, ISSNs, a Linking ISSN, and series information can be tagged. NISO STS makes it possible to describe multiple standards organizations as co-producers, record their full and abbreviated names, define a hierarchy of responsibility, and assign a role (e.g., "lead" or "co-producer" to each).

- **Sub-parts of Standards** — Can now take standard metadata.

- **Authorization** — An element in NISO STS holds the name of an Accrediting Organization and another element names the authorization that the agency named in the Accrediting Organization has authorized for *the current* standards document.

The precise meaning and potential use of some elements have been clarified/reinterpreted. At a minimum, if you have used the element as part of ISO STS, check the Tag Library and examples concerning the description and examples for the following elements: <std-ref>, <release-date>, <pub-date>, <version>, and <meta-date>.

2.9.2 OASIS Exchange (CALS) Tables

The Interchange versions of NISO STS use only the JATS XHTML-inspired table model, as defined in the NISO JATS XHTML-inspired Table Module. The Extended versions of NISO STS can use the OASIS XML Exchange (CALS) table model either instead of or in addition to the XHTML-inspired model.

To enable two different tagging schemes for the same material (row and column tables), all OASIS elements have been associated with a namespace and use the namespace prefix "oasis". (Since DTDs do not really work with namespaces, the namespace prefix has been hard-coded into the names of the OASIS table elements.) This differentiates the two sets of table elements, allowing both XHTML-inspired and OASIS XML Exchange (CALS) table models to be used together in a single standards document or by the same organization in different standards documents.

================