

Introduction to Multi-language Documents in NISO JATS

Deborah Aleyne Lapeyre[✉] and B. Tommie Usdin.

Mulberry Technologies, Inc.

[✉]Corresponding author.

In today’s multicultural world, many journal articles contain content in more than one language. The current NISO JATS Tag Sets provide several structures for encoding documents in which some of the metadata or text is provided in multiple languages. Multi-language content in NISO JATS can be handled using three simple techniques: 1) identification of the language using the `xml:lang` attribute, 2) repetition of some structures to enable these structures to be present in more than one language, and 3) enclosing some repeated structures in a single wrapper element, to indicate that they contain representations of a single logical object in different languages.

The latest NISO JATS Tag Sets, while not truly multilingual, provide rich support for multiple languages. The `xml:lang` attribute is practically ubiquitous, allowing most elements to state their language. Many metadata elements have been made repeatable, so that they can be present in the metadata once for each language. Specific NISO JATS elements also make it possible to encode an author’s name and affiliations in several languages or language/script combinations (without creating the false impression that these variations represent additional authors). This paper uses XML-tagged samples to illustrate these elements and others such as multi-language keywords and interleaved-language bibliographic citations.

The Issue of Multiple Languages

Most journal articles are published in one language, often but not always English. However, a significant number of articles contain content in multiple languages. We have observed several patterns of multi-lingual articles (and speculated about many others).

Within articles that are considered to be completely in *one* language, we often see:

- Names of people in multiple languages or scripts, where there may be several versions of the name of a single person;
- Names of institutions such as universities in multiple languages or scripts;
- The article title, journal title, or abstract “translated” into a second language;
- Foreign phrases embedded in the text, often displayed in italic; and
- Bibliographic references (citations) to materials published in a language other than the language of the paper, where information is tagged in *both* the primary language of the paper and the original language of the material being cited.

Within articles that clearly contain *multiple* languages, the pattern of multi-language usage varies considerably. Sometimes the languages are parallel and equal, sometimes the second language is considered a “translation” of an original. An article may include metadata in multiple languages, textual content in multiples, table and equation multiples, and/or multi-language citations. We have seen articles with:

- One primary language and selected metadata (such as titles, abstracts, and keywords) in additional language(s);
- One primary language and key textual content (such as references, tables, and figure captions) in a different language from the main text;

- One primary language and key content (such as references, tables, and equations) in both the language of the main text and duplicated in an additional language;
- All of their content in two or more languages, sometimes where the content is displayed in multiple columns on the page, each column containing a different language. This multi-column display may be an article and its translation(s), or it may be equivalent primary languages.

The NLM DTD, on which the NISO JATS Tag Sets were based, assumed that documents would be written and published in *one* language and was not designed for full multilingual processing. (From that NLM heritage, the default language for NISO JATS is English, but since JATS has been XML from day one and XML implies Unicode, NISO JATS has always been adequate for publishing articles in English, Korean, German, Russian, Japanese, or any other language modern Unicode could display, as long as users did not mind the tags being in English.) Shortly after the NLM DTD was published, it became clear that this Tag Set was meeting a significant need, but that many of its users did *not* publish in a mono-lingual environment. Tagging to handle multi-language material was gradually added to the Tag Sets by user request. While the methods NISO JATS now has for encoding multi-language material may not be as smooth as they might have been if multilingual material had been a primary requirement from the beginning, NISO JATS can accommodate common practice in multi-language journal articles.

Tagging Multiple Languages

Multi-language content in NISO JATS is handled through the use of three techniques:

1. Identification of the language of the document as a whole or of portions of the document (e.g., this article is in Norwegian; this abstract is in Greek);
2. Repetition of some structures (even components in the metadata that would never have multiple instantiations), to enable these structures to be present in more than one language (e.g., the same copyright statement can appear in both Polish and English); and
3. Enclosing some repeated structures in a single wrapper element to indicate that they contain representations of a single logical object (e.g., one university name expressed in three different languages).

This section explains and illustrates these three techniques. The remainder of this paper is an exploration of the three areas in a NISO JATS-tagged journal article where these techniques (in many variations) can be applied: the metadata (article, journal, and issue), the narrative content (body) of an article, and article back matter such as bibliographic reference lists.

Using `xml:lang` for Language Identification

The first and most obvious tool for marking up a multi-language document is the `xml:lang` attribute. NISO JATS did not invent this attribute; `xml:lang` is one of the few attributes described in the XML specification, which says:

In document processing, it is often useful to identify the natural or formal language in which the content is written. A special attribute named `xml:lang` may be inserted in documents to specify the language used in the contents and attribute values of any element in an XML document. In valid documents, this attribute, like any other, **MUST** be declared if it is used. The values of the attribute are language identifiers as defined by [IETF BCP 47], *Tags for the Identification of Languages*; in addition, the empty string may be specified [XML 1.0].

To see how `xml:lang` works, consider the simplest use of multiple languages in a NISO JATS document: tagging a word or phrase in a second language that is embedded in content written in a primary language. The phrase is tagged as NAMED CONTENT with an `xml:lang` attribute to indicate the language. For example:

```
... and on his left arm was tattooed <named-content
xml:lang="la" content-type="foreign-phrase">faber est suae
quisque fortunae</named-content> in bright red letters ...
```

or less colorfully:

```
Tonight we are dining <named-content xml:lang="fr"
content-type="foreign-phrase">en famille</named-content>.
```

This is our first technique; the attribute `xml:lang` is used to identify the language and/or script of a document or portion of a document.

The version 0.4 of NISO JATS greatly expands which elements can take `xml:lang` compared to previous versions of the JATS. The `xml:lang` attribute was added to *almost* all the elements in all three NISO JATS Tag Sets, so it is now much easier to list the elements that do *not* take `xml:lang` than to list those that do. The rare exceptions that do not take `xml:lang` include:

- High-level component container elements (such as FRONT, BODY, BACK, ARTICLE METADATA, and FLOATS GROUP);
- Elements that the Tag Sets do not control (such as all MathML and XHTML table elements); and
- Certain metadata elements such as the counts and identifiers, which should contain only numbers or character sequences and not text, per se.

Inheritance: Defaulting `xml:lang`

Unlike the typical attribute in an XML document, the value of `xml:lang` is inherited down the XML document tree hierarchy. If an element is in English, all of its descendents (its children and their children and their children ...) are expected to be in English, unless the language code is explicitly overridden. This means that a value for `xml:lang` given on the ARTICLE element is assumed to describe all content in the document unless there is another overriding `xml:lang` value on a component inside the document (such as a specific table or a foreign phrase). So, for example, if a document is identified as being in English (`xml:lang="en"`), it is assumed that all of the words in that document (whether in text, headings, equations, or tables) are in English, unless the document also contains, for example, an epigraph in Latin would be identified with `<disp-quote xml:lang="la">`.

In previous versions of the JATS Tag Sets, the default value of `xml:lang` was “en” (English) on the top-level element (ARTICLE), *and also* on some lower-level elements (such as the JOURNAL TITLE, JOURNAL SUBTITLE, ABBREVIATED JOURNAL TITLE, and SUB-ARTICLE). For articles in English this was redundant; for articles in other languages it required that the default be over-ridden multiple times, which is at best inconvenient and is likely to cause language encoding errors.

The current NISO JATS provides a default value for `xml:lang` only on the top level (ARTICLE) element, so that one over-ride at this point will cast the entire document into whatever language is specified. For example, `<article xml:lang="mo">` at the root of a document will indicate that the entire document is in Moldavian except any portions that have been specifically overridden to be in some other language (e.g., `<table-wrap xml:lang="de">`) indicates that a table is in German.

Recording Both Language and Script

Identification of language using `xml:lang` is not as simple as it might seem. That is, specifying a language is not always as easy as going to a lookup table of two-letter abbreviation language codes and choosing a code for the language (e.g., “ro” for “Romanian”). While some languages can be identified simply with one value, others have local variations (Canadian French is different in some significant ways from French as spoken in France), and a number of the world’s languages can be represented in more than one script.

NISO JATS uses `xml:lang` to record both the language and the script. For example, Japanese journal articles may include many scripts including Kanji, Hiragana, Katakana, Hiragana and Katakana combined, and “Japanese” (which is a combination of Han plus Hiragana plus Katakana). So a Japanese journal editor might need to mark a personal name as written in Kanji as opposed to Katakana (Kana), because “Kana names are frequently used to sort Kanji data”

[SPJ Working Group 2010].

In using `xml:lang` to hold both the language and the script codes, NISO JATS follows the W3C best practice guideline: *Network Working Group Request for Comments: 5646 Tags for Identifying Languages* [Phillips 2009].

That document defines a language tag as composed of (in part):

- A language code [called a subtag] (using the shortest ISO 639 code and possibly extended language subtags)
- Potentially followed by a hyphen and then
- A script code [called a subtag] (using an ISO 15924 code)
- Potentially followed by a hyphen and
- A region code [called a subtag] (also using an ISO 15924 code)

In *Appendix A* of that document (informative rather than normative), examples similar to the following are given to illustrate the use of language codes:

- Simple language subtags: `fr` (French), `ja` (Japanese), `zh` (Chinese)
- Simple script subtags:
 - `Hans` (Simplified variant)
 - `Hant` (Traditional variant)
- A Language subtag plus a Script subtag:
 - `zh-Hant` (Chinese written using the Traditional Chinese script)
 - `zh-Hans` (Chinese written using the Simplified Chinese script)
- A Language-Script-Region subtag combination:
 - `zh-Hans-CN` (Chinese written using the Simplified script as used in mainland China)
 - `sr-Latn-RS` (Serbian written using the Latin script as used in Serbia)

The Internet Assigned Number Authority (IANA) web site maintains the language-subtag registry, the lookup for language and script codes (<http://www.iana.org/assignments/language-subtag-registry>).

Repetition of Structures

Structures that are commonly published in multiple languages in the same document are allowed to repeat in the current version of the NISO JATS. While it is not required that repetitions of such elements use `xml:lang` attributes to identify the language of the repeated structure, this is highly recommended.

The `KEYWORDS GROUP` element is an example of a repeating structure, as the current NISO JATS allows `KEYWORD GROUPS` to repeat and to be identified by both a language and a keyword source. There may be multiple `KEYWORD GROUPS` as part of an `ARTICLE` for many reasons: there may be keywords provided by the authors, keywords provided by the publisher, and keywords selected from one or more controlled vocabularies. In addition, there may be `KEYWORD` elements in several languages, with `KEYWORD GROUP` repeated as many times as necessary, once per language:

```
<kwd-group xml:lang="en">
  <kwd>heated air</kwd>
</kwd-group>
```

```
<kwd-group xml:lang="ja">
  <kwd>加温空気</kwd>
</kwd-group>
```


This is our second technique: repeating structures, using `xml:lang` to mark the language of each structure.

Caveat: Since the language attribute is on the `KEYWORD GROUP` *rather* than on the individual `KEYWORD`, NISO JATS requires `KEYWORD` elements to be grouped by language/script and does not allow a `KEYWORD GROUP` to contain multiple languages.

Wrapped Repeating Structures Make One Object

It would be surprising in the world of journal articles, if anyone cared how many keywords an article had, or how many articles used a particular keyword (except, perhaps, as an indication of the popularity of specific topics of inquiry). However, there are parts of journal articles that are counted and where the statistics are of great importance. If a table is provided in both Korean and English, most readers will prefer to see only one or the other, and the count of tables in the article should include that table only once. More importantly, if an author name appears in two or three languages it is important that the author receive “credit” for only one publication, not for a publication in each of the versions of that author’s name.

Therefore, in some situations, it is important that all the repetitions of a component be identified as being equivalent (different names for the same component). If an equation is provided in MathML, TeX, and as a `.jpg` image, it is important it be:

- Counted as one equation (rather than three);
- Given a single number if numbers are programmatically generated; and
- Rendered once, using the most appropriate version of the equation for the device.

The mechanism to tie multiple versions of the same content into one countable/displayable set of options in NISO JATS is to wrap them in an `ALTERNATIVES` element. So the multiple versions of the equation just mentioned would be enclosed in a single `ALTERNATIVES` element to mark them as a single equation. Here is an equation in three forms: ASCII text with XML tags; a graphic; and as presentational MathML:

```
<disp-formula>
  <alternatives>
    <textual-form>(a + 3)<sup>2</sup> - (10 - b) = 24</textual-form>
    <graphic
      xmlns:xlink="http://www.w3.org/1999/xlink"
      xlink:href="fors2682.f3">
      <object-id>463492</object-id>
    </graphic>
    <mml:math>
      <mml:mrow><mml:msup><mml:mfenced>
        <mml:mrow><mml:mi>a</mml:mi><mml:mo>+</mml:mo>
        <mml:mn>3</mml:mn></mml:mrow>
      </mml:mfenced>...<mml:mn>24</mml:mn>
      </mml:mrow>
    </mml:math>
  </alternatives>
</disp-formula>
```

Similarly, multiple versions of a personal name can be enclosed in a `NAME ALTERNATIVES` element. Several alternatives of a single name might be tagged as:

```
<name-alternatives>
  <name specific-use="stage">
    <surname>Wayne</surname>
    <given-names>John</given-names>
  </name>
  <name specific-use="personal">
    <surname>Morrison</surname>
    <given-names>Marion Mitchell</given-names>
```

```
</name>
<name specific-use="birth">
  <surname>Morrison</surname>
  <given-names>Marion Robert</given-names>
</name>
</name-alternatives>
```

and another tagged as:

```
<name-alternatives>
  <name name-style="eastern" xml:lang="ja-Jpan">
    <surname>天海</surname>
    <given-names>祐希</given-names>
  </name>
  <name name-style="eastern" xml:lang="en">
    <surname>Yūki</surname>
    <given-names>Amami</given-names>
  </name>
</name-alternatives>
```

A table in multiple formats (such as both a graphic and an XML-tagged XHTML table) would be handled similarly, with a single title and caption but the tabular content tagged as ALTERNATIVES.

This is our third technique: repeating structures within a wrapper element to indicate that there are not multiple structures, just variants of a single real-world object.

Details: Metadata in Multiple Languages

To allow metadata in more than one language, many of the NISO JATS metadata elements that contain prose can be repeated as many times as necessary, with an `xml:lang` attribute indicating the language of each repeated metadata item. Some elements, for example, ABSTRACT and ALTERNATIVE TITLE have always been modeled as repeating. Some elements, for example, small textual elements such as ISSUE TITLE, COPYRIGHT STATEMENT, and CONFERENCE NAME are now allowed to repeat, so that they can be provided in multiple languages. Some of the grouping elements (such as KEYWORD GROUP, CONFERENCE METADATA, and FUNDING GROUP) can also repeat, where each repetition of the complete structure allows for all the contents to be recorded in another language.

Abstracts

There is a choice to make when tagging abstracts provided in multiple languages. NISO JATS provides two elements for abstracts: ABSTRACT and TRANSLATED ABSTRACT, both of which may repeat and both of which can take an `xml:lang` attribute. Two methods are provided because they provide *subtly* different information:

- The repeating ABSTRACT element is for abstracts in the original language(s) of the article. Many articles will contain only one primary ABSTRACT, but an article that was originally published in two or three languages (both French and English, for example) could appropriately contain two or three primary ABSTRACT elements that differ only in language.
- The TRANSLATED ABSTRACT element is for abstracts in languages other than the original language of the article. This is used for the English translation of an abstract that was originally published in Finnish or Norwegian.

Journal Metadata

The JOURNAL METADATA element is a place where repetition has been deliberately introduced during the most recent NISO JATS update to provide for metadata components in more than one language. The JOURNAL METADATA wrapper element itself does not repeat, but each of these elements inside JOURNAL METADATA may now repeat, and each can take a language attribute to differentiate among those repetitions: JOURNAL IDENTIFIER, the entire JOURNAL TITLE GROUP, JOURNAL TITLE, JOURNAL SUBTITLE, ABBREVIATED JOURNAL TITLE, and PUBLISHER NAME and PUBLISHER LOCATION (which repeat as a pair inside the PUBLISHER element).

As in the case of ABSTRACT elements, there are also choices to make when tagging multiple journal titles. NISO JATS provides, within the JOURNAL TITLE GROUP, four repeatable elements for titles: JOURNAL TITLE, JOURNAL SUBTITLE, ABBREVIATED JOURNAL TITLE, and a TRANSLATED TITLE GROUP (which contains a TRANSLATED TITLE and TRANSLATED SUBTITLE).

For tagging the other titles, there are choices, as there were in ABSTRACT and TRANSLATED ABSTRACT:

- The repeating JOURNAL TITLE is for titles in the original language(s) of the ARTICLE. Most articles will contain only one JOURNAL TITLE, but an article that was originally published in two or three languages could appropriately contain two or three primary JOURNAL TITLES. (Note: Another way to handle this is for the whole JOURNAL TITLE GROUP to repeat, once in each primary language. NISO JATS, as a conversion target, often allows multiple ways to encode the same information.)
- The repeating TRANSLATED TITLE GROUP binds together a title and an optional subtitle in languages other than the original language of the article. This grouping of TRANSLATED TITLE and TRANSLATED SUBTITLE(s) prevents confusion as to which subtitle should be paired with which title.

Article Titles

The structures for handling multi-language article and journal titles in the NISO JATS pre-date the effort to make the major structures of the model multi-lingual, and thus they are shaped a bit differently from the newer and more flexible multi-language structures.

The NISO JATS structure assumes that there is one and only one title for an article, which uses the element ARTICLE TITLE. This unitary ARTICLE TITLE model assumes a distinction between primary titles and secondary “translated” titles. There may be as many ARTICLE SUBTITLE(s) elements associated with an ARTICLE TITLE as necessary. It is assumed that the ARTICLE TITLE and the ARTICLE SUBTITLE(s) will be in the same language.

Article titles in additional languages are wrapped in TRANSLATED TITLE GROUP elements. Each TRANSLATED TITLE GROUP may take an `xml:lang` (which is highly recommended, since a translation into an unspecified language is unlikely to be useful) and contains a TRANSLATED TITLE and as many TRANSLATED SUBTITLE elements as necessary. This grouping of TRANSLATED TITLE and TRANSLATED SUBTITLE(s) prevents confusion as to which subtitle should be paired with which title.

Note: It might have been a cleaner solution to tag two separate and equal ARTICLE TITLE elements, differentiated by a language attribute, in the way that multi-language ABSTRACT elements are tagged, but at the time of the design, there were the following concerns:

- It was felt that most journal articles are written in *one* language, even if they are then published simultaneously in three (this may not be *as* true of laws and regulations, books, or other publication forms);
- It was felt that the ARTICLE TITLE (unlike a repeating ABSTRACT or even a repeating JOURNAL TITLE) is of primary importance in the indexing and discoverability of an article; and
- The TRANSLATED TITLE and TRANSLATED SUBTITLE elements were to be preserved for backwards compatibility purposes.

Names of People

Flexibility of the Basic Personal Name Model

The NISO JATS Tag Set can encode many, if not most, of the name variations found in the published journal articles. In NISO JATS, the NAME element is the container for the component elements of personal names, such as SURNAME and GIVEN NAMES. The (brand new) model for the NAME of a person is:

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

This model provides a great deal of flexibility to deal with many of the world’s naming conventions by allowing a personal name to be:

- Only a SURNAME element (used for westernized single names such as “Pele”, “Aztek” or “Ice Cube”),
- Only a GIVEN NAMES element (For Tibetan, Indian, and Burmese names that are single names *but not surnames*, there is no need to commit the tag abuse of calling this name a SURNAME, the GIVEN NAMES element can stand alone.), or
- A SURNAME followed by a GIVEN NAMES (the naming pattern typical in American, English, and many European names).

Thus, all of the following are valid NISO JATS Names:

```
<name>
  <given-names>Cai-Rang</given-names>
</name>

<name name-style="eastern">
  <surname>Chou</surname>
  <given-names>Hsun-Tse</given-names>
</name>

<name name-style="western">
  <surname>Smith</surname>
  <given-names>Jane Alexandra</given-names>
</name>

<name>
  <surname>Cher</surname>
</name>
```

Since it is very important in many cultures, both the GIVEN NAMES and SURNAME elements may contain multiple words:

```
<surname initials="Q">Llanos De La Torre Quiralte</surname>
<given-names initials="M">M</given-names>

<surname initials="L">Lapeyre</surname>
<given-names initials="KPC">Kenneth Pritchard Carnu</given-names>
```

When tagging a personal name in NISO JATS, there is no need to separate given names into first and middle, which does ease one editorial burden. This separation can be problematic even in English with combined first names like “Dottie Jean” and “Mary Dawn”:

```
<surname initials="W">Williams</surname>
<given-names initials="DG">Dottie Jean Gray</given-names>

<surname initials="A">Austin</surname>
<given-names initials="M">Mary Dawn</given-names>
```

and also in determining whether an opening initial counts as a first name:

```
<surname initials="U">Usdin</surname>
<given-names initials="BT">B. Tommie</given-names>

<surname initials="S">Sperberg-McQueen</surname>
<given-names initials="CM">C. Michael</given-names>
```

NISO JATS personal name tagging can be quite flexible. Parts of a name, such as “de” and “de la”, can be tagged in place as they fall within the name, positioned at the front of a name, or relegated to the rear of the surname following a comma:


```
<surname>Llanos De La Torre Quiralte</surname>
<given-names>M</given-names>
```

```
<surname>Toulouse-Lautrec-Monfa, de</surname>
<given-names>Henri Marie Raymond</given-names>
```

How Language May Affect a Name

Care needs to be taken with multipart names to divide the components into family names (SURNAME) and personal names (GIVEN NAMES) in a culturally appropriate fashion. The finer points of personal names should be determined by native speakers of the specific language in geographically appropriate ways. The following examples illustrate a small bit of the complexity in language-specific naming and show how NISO JATS tagging can be adapted to different naming situations, *given sufficient editorial knowledge*.

Spanish. The traditional order for Spanish names is first-name/middle-name/first-surname/second-surname, usually indexed by the first surname, which is typically the patronymic. Since “western” sort order sorts by SURNAME, this style fits NISO JATS modeling well and can be tagged by:

- Placing the first-name and middle names inside the GIVEN NAMES element;
- Placing both surnames in the proper order inside the SURNAME element; and
- Setting the NAME STYLE attribute to “western”.

Portuguese. Although Portuguese names may appear similar to Spanish names to a non-native speaker, tagging a Portuguese name is more complex from the NISO JATS perspective. The traditional order for Portuguese names is first-name/middle-name/first-surname(matronymic)/second-surname(patronymic), indexed by the second surname (patronymic) [Black 2003]. If both surnames are placed within the SURNAME element, as would be done for a Spanish name, the sort and indexing orders will not be correct. Several potential compromises are possible (none completely satisfactory and all verging on tag abuse), including:

- Tagging a single name by placing both surnames inside the SURNAME element: the patronymic, followed by a comma, followed by the matronymic and setting the NAME STYLE attribute to “western”.
- Creating two names inside a Names Alternatives wrapper. In one of these (the primary name,) recording the full (matronymic then patronymic) surname. In the other alternative NAME, placing the first-name, middle name, and matronymic inside the GIVEN-NAMES element; the patronymic inside the SURNAME element; and set the order to “western”. Use the SPECIFIC USE attribute on this second name to describe a use such as sorting or indexing to help prohibit abuse.
- Placing the entire name in a STRING NAME, tag the GIVEN NAMES, and tag only the patronymic as a SURNAME, leaving the matronymic as untagged text.

String Name

The STRING NAME is a very loose element, which may contain text, numbers, special characters, generated text, and any or all of the naming elements, such as SURNAME and GIVEN NAMES. How much to tag within a STRING NAME must be determined editorially and validated with an external mechanism such as Schematron, since the DTD provides no structure. A STRING NAME can be used to hold: a sort version of a name (for example with accented characters replaced with lower-ASCII letters), a full name in display order (for example, as a byline that does not require recombining name components), name components with unusual punctuation or spacing between them, or the text of a name for which the given-names/family-name distinction does not exist or cannot be determined.

Here is a fully-tagged STRING NAME:

```
<string-name>
  <prefix>The Honorable</prefix> <given-names>John Mesach
  Irving Browning</given-names> <surname>Jones</surname>,
```

```
<suffix>III</suffix>
</string-name>
```

a partially tagged STRING NAME:

```
<string-name>The Honorable <given-names>John Mesach Irving
  Browning</given-names> <surname>Jones</surname>, III
</string-name>
```

and an untagged STRING NAME:

```
<string-name>Prince Charles</string-name>
```

Thus STRING NAME can be a mechanism for providing a *display* version for less regularly-constructed names such as Arabic names. Traditional Arabic names (with some variation by tribe, region, and country) do not follow the western style of first-names/middle-names/last-names; rather they contain (potentially) 5 parts: “ism”, “kunya”, “nasab”, “laqab”, and “nisbah”, arranged in varying order. The “ism” is (usually and only approximately) the given name. The “kunya” is an honorific not typically printed or displayed. Both the “laqab” and the “nisbah” can be used as surnames but are not in all cases. The “nasab” is frequently, but not always, the patronymic [Notzon and Nesom 2005].

In the NISO NISO JATS *Archiving* Tag Set, STRING NAME may be used everywhere that NAME is allowed, but in the NISO JATS *Publishing* Tag Set, STRING NAME is only allowed within NAME ALTERNATIVES.

Name Display Order

NISO JATS name tagging also handles the more basic cases of name display order. The NAME STYLE attribute records the preferred display order for a name, for example between Eastern and Western display order (“Toshiro Mifune” versus “Mifune Toshiro”). NAME STYLE information can be used for choosing an inversion algorithm for sorting, for ordering the names for display, or for other processing functions. The attribute NAME STYLE can be used on both NAME and STRING NAME so that “eastern”, “western”, or “islensk” sort and display order can be specified, even for a name with many untagged components.

The values of the NAME STYLE attribute and their approximate meanings are given below.

- When the value is “western”
 - The display order is: given (GIVEN NAMES) followed by family (SURNAME); and
 - The sort/inversion order is family (SURNAME) then given (GIVEN NAMES).
- When the value is “eastern”
 - The display order is: family (SURNAME) followed by given (GIVEN NAMES); and
 - The sort/inversion order is family (SURNAME) then given (GIVEN NAMES).
- When the value is “islensk”
 - The display order is: given (GIVEN NAMES) followed by patronymic (SURNAME); and
 - The sort/inversion order is given (GIVEN NAMES) then patronymic (SURNAME).

Here are two samples of NAME STYLE in use:

```
<name name-style="western">
  <surname>Jones</surname>
  <given-names>John Mesach Irving Browning</given-names>
  <prefix>The Honorable</prefix><suffix>III</suffix>
</name>
```

```
<name name-style="eastern">
  <surname>Chou</surname>
  <given-names>Hsun-Tse</given-names>
```

</name>

Note: Whether a name is “eastern” or “western” in style does not refer to the characteristics of the name, but to the desired display and sort order. For example, both of the following are valid names that reflect the preference of the author for the family-name-followed-by-given-names display order:

```
<name name-style="eastern" xml:lang="en">
  <surname>Nakanishi</surname>
  <given-names>Hidehiko</given-names>
</name>
```

```
<name name-style="eastern" xml:lang="ja-Kana">
  <surname>ナカニシ</surname>
  <given-names>ヒデヒコ</given-names>
</name>
```

Note: Perhaps a western bias has not left NISO JATS entirely, because the default value for the attribute Name Style is “western”.

NAME ALTERNATIVES: More than One Name for an Individual

As part of the considerable effort toward internationalization and multiple language support that was undertaken for the latest NISO JATS, a new wrapper element was added to hold more than one version of a personal name. There is no limit to the number of alternatives that may express a single name; thus it is possible to name the same person in, for example, Hiragana, Katakana, and a Romanized form in the Latin alphabet.

The NAME ALTERNATIVES element can be used to record:

- Alternative versions of a name in multiple native languages and multiple scripts (as has already been shown);
- Versions of a name with and without ligatures and diacritics so that display can show an “é” and indexing, sorting, or searching can revert to an accentless lower-ASCII “e”.
- Transliterated versions of an name;
- Vancouver-style abbreviated names (Smith AK, Jones BC, Bloggs TC); or
- Additional names for indexing (For example, it may be desirable to record in an XML database all the name variants for an individual, from “President Thomas Jefferson” to “Long Tom”, with the attribute SPECIFIC USE marking “primary” versus “index”).

It is critical that multiple names be clearly tagged as *alternatives* to avoid the implication of multiple authors when more than one form of a name is given. The new NAME ALTERNATIVES element groups multiple versions of the name of one person in the same way that the ALTERNATIVES element groups processing alternatives for one graphic or one table. These multiple names *are* processing alternatives, since an application must choose whether to display or otherwise use only one of the names or more than one.

The content of the NAME ALTERNATIVES element is one or more NAME and/or STRING NAME elements. Note: This new element thus introduced STRING NAME into the NISO JATS *Publishing* Tag Set, where it had not previously been allowed. Within the Publishing Tag Set model, STRING NAME is permitted *only* as an alternative. The Tag Library documentation for Publishing states that a STRING NAME should not be used for the primary name, which should still be tagged in strict order with a SURNAME element, followed by a GIVEN NAMES element, et al. NAME ALTERNATIVES should be used in Publishing only to support language variants or indexing and searching alternatives.

The NAME ALTERNATIVES element is allowed in all the places NAME is allowed, with the typical usage expected to be inside a CONTRIBUTOR element. (Note: Therefore, for both Archiving and Publishing Tag Sets, the new element can occur inside: CONTRIBUTOR, ELEMENT CITATION, MIXED CITATION, PERSON GROUP, PRINCIPAL AWARD RECIPIENT, PRINCIPAL INVESTIGATOR, PRODUCT, RELATED ARTICLE, and RELATED OBJECT.)

Here are three alternatives, showing the same name in long display form, with parts identified, and abbreviated

display form:

```
<name-alternatives>
  <string-name
    specific-use="display">José del Pogo García</string-name>
  <name name-style="eastern">
    <surname>del Pozo García</surname>
    <given-names>José</given-names>
  </name>
  <string-name
    specific-use="abbrev-form">Pozo Garcia J del</string-name>
</name-alternatives>
```

Here are three alternatives, showing the same name in three language/script variants:

```
<name-alternatives>
  <name name-style="eastern" xml:lang="ja-Jpan">
    <surname>中西</surname>
    <given-names>秀彦</given-names>
  </name>
  <name name-style="eastern" xml:lang="en">
    <surname>Nakanishi</surname>
    <given-names>Hidehiko</given-names>
  </name>
  <name name-style="eastern" xml:lang="ja-Kana">
    <surname>ナカニシ</surname>
    <given-names>ヒデヒコ</given-names>
  </name>
</name-alternatives>
```

Here are two STRING NAME alternatives, showing the same name in display form and abbreviated form:

```
<name-alternatives>
  <string-name
    specific-use="display">PM Sudha</string-name>
  <name><given-names initial="PM">Sudha</given-names>
  </name>
  <string-name
    specific-use="abbrev-form">Sudha PM</string-name>
</name-alternatives>
```

Here are three alternatives, showing the same name as a formal name, as a common name, and as typically abbreviated.

```
<name-alternatives>
  <name content-type="formal-name" xml:lang="fr">
    <surname>Giscard d'Estaing</surname>
    <given-names>Valéry Marie René Georges</given-names>
  </name>
  <name content-type="common-name" xml:lang="fr">
    <surname>Giscard d'Estaing</surname>
    <given-names>Valéry</given-names>
  </name>
  <string-name
    specific-use="abbrev-form">Giscard d'Estaing V</string-name>
</name-alternatives>
```

Affiliation/Institution Names

Sometimes an affiliation needs to be recorded in multiple languages or multiple scripts, for example, the name of a contributor’s institute or university once in English and once in German (both “Institute for the German Language

(IDS)” and “Institut für Deutsche Sprache”). When repeating the university name, care must be taken so that it does not appear to multiply the number of a contributor’s affiliations. Three versions of the name of a university is not the same as three different universities!

Affiliation Alternatives

The NISO JATS `affiliation alternatives` element can record `affiliation` names (such as universities) in more than one language, by collecting together all the representations of a single `affiliation`; the `xml:lang` attribute can be used to distinguish the different `affiliation` elements for separate processing. It will be up to an application to determine how multiple versions of a single affiliation are processed. In the example below, the `xml:lang` is placed on the institution; if it were more appropriate, the `xml:lang` attribute could be placed on the `affiliation` element.

```
<aff-alternatives id="aff1">
  <aff>
    <institution xml:lang="en">Institute for the German
      Language (IDS) </institution>
  </aff>
  <aff>
    <institution xml:lang="de">Institut für Deutsche
      Sprache </institution>
  </aff>
</aff-alternatives>
```

The `AFFILIATION ALTERNATIVES` element might be used to record:

- The name of an affiliation in multiple languages (for example, a university name in Korean or Chinese-Han characters and a transliterated version of the same name in the Latin alphabet);
- The name of an affiliation in multiple script combinations for a single language (for example, a university name in Japanese [`xml:lang="ja-Jpan"` for Han + Hiragana + Katakana] and the same university name written in Kanji [`xml:lang="ja-Hani"`]);
- An alternate affiliation for sorting or searching (for example, a name in French with accented letters (such as an “é”) and a plain-letter lower-ASCII version of the same name with “é” replaced by “e” for sorting. (Note: The the `SPECIFIC USE` attribute can be used to indicate that the ASCII version is only for “sorting”, not for display.)

Using AFFILIATION ALTERNATIVES

To tag several language-variant institutions, the `INSTITUTION` element inside an `AFFILIATION` does *not* repeat, but the entire `AFFILIATION` element does, taking an `xml:lang` attribute to identify the language of the `AFFILIATION` variant. Repeating the enclosing element (`AFFILIATION`) rather than the institution name (`INSTITUTION`) allows any elements within the `AFFILIATION`, such as `ADDRESS` or `COUNTRY`, also to be provided in an alternative language.

The `AFFILIATION ALTERNATIVE` element can be used anywhere that `AFFILIATION` is allowed. In the NISO JATS Publishing Tag Set, this means that the elements `AFFILIATION` and `AFFILIATION ALTERNATIVES` are allowed inside `ARTICLE METADATA`, `COLLABORATION`, `CONTRIBUTOR`, `CONTRIBUTOR GROUP`, `FRONT STUB` (for `SUB-ARTICLES`) and `PERSON GROUP` (inside citations). In Archiving, the elements `AFFILIATION` and `AFFILIATION ALTERNATIVES` may be in all of these places as well as inside `SIGNATURE BLOCK` elements.

Linking a Contributor to an Affiliation

Many, if not most, NISO JATS-based tag sets tie each contributor to his or her affiliation(s) using the XML `ID/IDREF` mechanism of XML. In a mono-language document, an `ID`-type (named “`id`”) attribute is placed on each `AFFILIATION` element

```
<aff id="xyz-1">...</aff>
```

and a cross-reference element (<xref>) inside a CONTRIBUTOR element points to this ID, using an IDREF.

```
<xref ref-type="aff" rid="xyz-1"/>).
```

When AFFILIATION ALTERNATIVES are tagged, it is best practice to place the ID-type attribute (that for a single affiliation would have been placed on the AFFILIATION element) onto the AFFILIATION ALTERNATIVES wrapper element instead. There is a single ID-type attribute, because there is only one institution, and putting the ID on the wrapper makes it clear that this is a connection between a contributor and a single affiliation, however many names are provided for that institution:

```
<aff-alternatives id="aff1">
  <aff xml:lang="en"> ... Affiliation in English...</aff>
  <aff xml:lang="sv"> ... Affiliation in Swedish...</aff>
</aff-alternatives>
```

Some publishers have chosen a more complicated and less straight-forward technique and place an ID on *only one* of the alternatives, thus marking it as the preferred display alternative while still providing language variants:

```
<aff-alternatives>
  <aff id="aff3">
    <institution xml:lang="ja-Jpan">国立言語学博物館</institution>
  </aff>
  <aff>
    <institution xml:lang="en">National Museum of
      Linguistics</institution>
  </aff>
</aff-alternatives>
```

Details: Body Content in Multiple Languages

Key Content in a Different Language from the Main Text

In some journals (for example, many of the scientific journals published in Korea), the prose of the article is in the local language (Korean) but metadata and selected content — considered key for international recognition and discoverability — are published in English. English structures include tables, figure captions, and bibliographic references.

Because these structures are not duplicated (they are published only in English, not in both English and Korean), the structures are simply identified in the XML as being in English with an `xml:lang` value of “en”. Other than that, there is nothing unusual about the way these structures are handled in the XML.

Key Content is Duplicated (Once per Language)

Providing some content in English seems to be common in journals that are published in other languages. In many cases, the content appears in the local language of the journal and is replicated in English. In these cases, how the duplicate structures are handled varies; publisher practices in this are not standardized.

Graphics and Media

Graphics and media objects are non-XML objects, typically a binary format such as .png, .jpg, or one of the video formats. In spite of that, there may be writing displayed in a picture, or the speech in a video may be in a particular language. Thus, both the GRAPHIC and MEDIA elements are cases of the wrapped-repetition solution, repeated inside an ALTERNATIVES wrapper and given an `xml:lang` attribute to differentiate.

In this example, we are trusting that the “classic” images are the same, differing only in internal language:

```
<graphic
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:href="http://www.zeed.org/publicimages/classic-en.jpg"
```

```
    xml:lang="en">
</graphic>
<graphic
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:href="http://www.zeed.org/publicimages/classic-fr.jpg"
  xml:lang="fr">
</graphic>
```

Tables

In NISO JATS, if a table (a structure of rows and columns) is provided in several forms or languages, all of the versions of the tabular material are wrapped in an `ALTERNATIVES` element inside a `TABLE WRAP` element. `ALTERNATIVES` may include structural tables in several languages, as well as format variants for table display such as `GRAPHICS` and `PREFORMATTED TEXT` versions of the table.

For example, a `TABLE WRAP` element might contain:

- A `CAPTION` (which does not repeat, but can take an `xml:lang` attribute, as can the paragraphs inside it);
- An `ALTERNATIVES` element containing all of the following options:
 - `TABLE` (`xml:lang="ru"` for an XHTML structural table in Russian)
 - `TABLE` (`xml:lang="en"` for an XHTML structural table in English)
 - `PREFORMATTED TEXT` (`xml:lang="en"` for a preformatted text of the table, in English)
 - `GRAPHIC` (`xml:lang="ru"` for a graphic representation of the table in Russian)
 - `GRAPHIC` (`xml:lang="en"` for a graphic representation of the table in English)

Article Body in Multiple Languages

There are several solid solutions in use (and many more proposed) to handle documents that exist in many languages simultaneously (for example, the laws of the European Union). Many Canadian documents do not name a *primary* language; the copy of record exists simultaneously in both English and French. NISO JATS was *not* designed to cope with true multilingual documents such as these.

However, a useful solution to the problem of a complete article body in two or more languages (both French and English) can be found in the `SUB-ARTICLE` element. Each language is treated as a separate `SUB-ARTICLE` under the master `ARTICLE`, one `SUB-ARTICLE` for each language. This works well for the French/English Canadian examples we have seen (mainly editorials and short features), in which all the figures and tables are placed in the common `FLOATS GROUP` element at the end of the main article. The main article need contain no text, only metadata, with all narrative content in the `SUB-ARTICLES`. This works also for the two-column multi-language examples (one column in English and the parallel column in Japanese) we have seen that share a single set of figures and tables.

One advantage of this technique is that this is not just a repeated `ARTICLE BODY` with a language attribute; each `SUB-ARTICLE` *may* include a full set of language-specific metadata. Search results in either language bring a reader to the single `SUB-ARTICLE`. Each `SUB-ARTICLE` contains its own references and, if these figures and tables are *not* shared, its own mono-lingual figures and tables.

Details: Back Matter in Multiple Languages

Back Matter Structures (repeated structure)

Nearly all the components of NISO JATS `BACK MATTER` can repeat and take a language attribute, so any individual component can be in a different language from the primary article. The following all repeat (with an `xml:lang` attribute): `ACKNOWLEDGMENTS`, both `APPENDIX GROUP` and the `APPENDIX(es)` inside it, `BIOGRAPHY`, `FOOTNOTE GROUP`, `GLOSSARY LIST`, `REFERENCE LIST (Bibliography)`, `NOTES`, and `SECTION`.

Multi-language References

Bibliographic references are considered to be critical aspects of an article, necessary for citation indexing, web links, and database lookup. In a multilingual world, there are many ways in which languages can be a factor in bibliographic reference lists. In some reference lists we find a mix of two or more languages, so there is a requirement to specify both the language in which the article was published and the language in which the citation is presented (which may not be the same).

Multiple Language References in a Single Reference List

Even within a single-language article, there may be bibliographic references in languages other than the primary language of the article. This is especially common if the material being cited is in a language other than that of the article.

If the entire REFERENCE element is in German (`xml:lang="de"`) in an otherwise English REFERENCE LIST, no special NISO JATS handling is required. The full REFERENCE or the citations inside it (MIXED CITATION and ELEMENT CITATION) may take the `xml:lang` attribute "de" to declare their language. Note that this `xml:lang` records the language of the *citation*, not necessarily the language of the *cited work*.

Here are two valid references as they might appear in an almost-all-English References List:

```
<ref id="r27">
  <label>27) </label>
  <mixed-citation publication-type="journal" xml:lang="ja">
    <person-group person-group-type="author">
      <name><surname>#x4F0A;#x85E4;</surname>
        <given-names>#x96C5;#x548C;</given-names></name>
      <name><surname>#x5CA1;#x672C;</surname>
        <given-names>#x771F;#x7531;#x7F8E;</given-names></name>
      <name><surname>#x9234;#x6728;</surname>
        <given-names>#x5553;#x4E09;</given-names></name>
      <name><surname>#x67F4;#x7530;</surname>
        <given-names>#x96C5;#x535A;</given-names></name>
      <name><surname>#x4F50;#x3005;#x6728;</surname>
        <given-names>#x5EB7;#x96C4;</given-names></name>
    </person-group>, <source>#x65E5;#x672C;#x539F;#x5B50;
    #x529B;#x5B66;#x4F1A;#x8A8C;</source>#x3001;
    <volume>36</volume>[<issue>11</issue>],
    <fpage>1055</fpage>&ndash;<lpage>1058</lpage>
    (<year>1994</year>).
  </mixed-citation>
</ref>

<ref id="r29">
  <label>28) </label>
  <mixed-citation publication-type="journal">
    <string-name><surname>Weissert</surname>,
    <given-names>W</given-names></string-name>,
    <string-name><surname>Wan</surname>,
    <given-names>T</given-names></string-name>,
    <string-name><surname>Livieratos</surname>,
    <given-names>B</given-names></string-name>,
    <string-name><surname>Katz</surname>,
    <given-names>S</given-names></string-name>. <article-title>
    Effects and costs of day-care services for the chronically ill: a
    randomized experiment</article-title>. <source>Medical
    Care</source> <year>1980</year>;<volume>18</volume>:
    <fpage>567</fpage>&ndash;<lpage>584</lpage>. [<pub-id
    pub-id-type="pmid">6772889</pub-id>].</mixed-citation>
</ref>
```

Multiple Single-language Reference Lists

In other multi-language documents, the entire Reference List may be provided in two or more languages. This also requires no special structures or handling in NISO JATS: the REFERENCE LIST element is allowed to repeat, and each REFERENCE LIST takes an `xml:lang` attribute. Some publishers prohibit this practice because they fear it could artificially double the number of citations.

Citations in Multiple Languages

In some multi-language documents, a single citation or parts of the citation may be provided in two or more languages. Great care must be taken to tag a multi-language citation as *one* citation so that it will be seen as a *single* citation and referenced from within the document as a single citation identifier. Because one REFERENCE may legitimately contain more than one citation (MIXED CITATION OR ELEMENT CITATION), just repeating the citation with an `xml:lang` attribute is not an adequate solution. Such content would be properly interpreted as two citations inside a single REFERENCE.

A study of multi-language citations in many journals showed that the style for publishing such references varies from one publisher to another.

- Some publishers cluster all the elements of one language together (a full citation in French followed by the same full citation in English);
- Some publishers intersperse single-language clusters of a second language inside a citation that is largely in a single primary language; and
- Some publishers alternate languages (ARTICLE TITLE in English followed by ARTICLE TITLE in Japanese; SOURCE in English followed by SOURCE in Japanese).

NISO JATS enables many techniques for making multi-language references, but does not prescribe any one technique, thus allowing a publisher to determine the mixture. NISO JATS tagging copes with the great variability by marking individual elements within a citation (ARTICLE TITLE, SOURCE, PUBLISHER, etc.) with `xml:lang` to indicate those elements in a language other than the primary language of the document. When the two languages are sequential (all Japanese elements followed by all English elements) this solution has the less-than-optimal property that there is no “wrapper” element around all the parts of a single language. Since all other styles of interspersed languages within a reference were resolved by the `xml:lang` solution, perhaps this infelicity is more philosophical than practical. (If a publisher or archive were so inclined, they could delineate the first part of a citation as in language one and the second part as in language two, and use milestone elements to mark the two parts. We do not recommend this approach.)

Here is a citation with several components marked with languages:

```
<ref>
  <mixed-citation publication-type="journal"
    publication-format="print">
    <string-name>
      <surname>Llanos De La Torre Quiralte</surname>
      <given-names>M</given-names></string-name>,
      <string-name>
        <surname>Garijo Ayestaran</surname> <given-names>M
        </given-names></string-name>, <string-name>
        <surname>Poch Olive</surname> <given-names>ML
        </given-names></string-name>.
      <article-title xml:lang="es">Evolucion de la mortalidad
        infantil de La Rioja (1980-1998)</article-title> [<trans-title
        xml:lang="en">Evolution of the infant mortality rate
        in la Rioja in Spain (1980-1998)</trans-title>].
      <source xml:lang="es">An Esp Pediatr</source>. <year>2001</year>
      <month>Nov</month>;<volume>55</volume>(<issue>5</issue>):
      <fpage>413</fpage>-<lpage>420</lpage>. Figura 3, Tendencia
      de mortalidad infantil [Figure 3, Trends in infant
      mortality]; p. 418. Spanish.</mixed-citation>
</ref>
```

For creating multi-language references, NISO JATS Tag Libraries make a number of Best Practice recommendations, which we have expanded below:

- A contributor name in more than one language should be placed inside a `NAME ALTERNATIVES` grouping element. This ensures that one author does not get cited multiple times for a single article. All the named alternatives in such a grouping represent the same author.
- An `ARTICLE TITLE` in more than one language may either use the `TRANSLATED TITLE` element for the second title, or use the element `ARTICLE TITLE` for both titles, with an `xml:lang` attribute to differentiate the two. When there are two primary languages, best practice is to tag two separate and equal `ARTICLE TITLE` elements, differentiated by a language attribute. With a primary title and an obvious translation (for example, an `ARTICLE TITLE` in Korean and the transliterated Romanized title in English) the `TRANSLATED TITLE` element could be used as an alternative. Which of these is best practice depends on editorial design and the ability of expected search services (e.g., When a title search is performed, does the search engine just hunt through `ARTICLE TITLE` elements or also through `TRANSLATED TITLE` elements?).
- A `SOURCE` (such as a book title or abbreviated journal name) in more than one language may either use the `TRANSLATED SOURCE` element for the second source, or use the element `SOURCE` for both sources, with an `xml:lang` attribute. (Note: It is probably a slightly cleaner solution to tag two separate and equal `SOURCE` elements, differentiated by a language attribute.)
- Other elements within a citation that are present in more than one language may use the `xml:lang` attribute on each element to mark the language or just mark those in the second (and all subsequent) languages.

Language in which the Cited Article Was Published

NISO JATS has no special element to indicate the language in which a cited article was originally written, and the issue of where and how that is said within a citation (if it is said at all) can be complex. In most journal citations, if a language attribute is set on the `ARTICLE TITLE`, that can be inferred to be the original language of the article. However, in some journals, particularly those in chemistry and physics, the `ARTICLE TITLE` is not displayed in the citation and may not be present, even in the electronic record. In this situation, the `xml:lang` attribute can be placed on the `SOURCE` element, which names the journal or book in which the article was published. Again, the inference can reasonably be made that the `SOURCE` is in the language of the article, but some European periodicals publish in multiple languages, so this might be misleading.

It is possible to place `xml:lang` on all of the elements inside a citation, and that might be definitive. But many journals publish this material *in the narrative text* of the citation (e.g., the words “(in Japanese)” or “translated from the Russian”). In `MIXED CITATIONS`, such phrases may just be text; in Element Citations, the material can be tagged as a `COMMENT` element with a `CONTENT TYPE` attribute indicating that this is language information (for example, “language”, “language-info”, or “original-language”).

Acknowledgments

We would like to thank the members of the Scholarly Publishing in Japanese (SPJ) Working Group, who both enhanced our personal knowledge concerning the nature of the many multi-language issues and, most helpfully, provided many Japanese XML examples. どうもありがとうございました。(Thank you very much.)

Thanks also to Felix Sasaki, who provided expertise on internationalization, Japanese naming, and Japanese typography.

Additional thanks to the publication *Science Editor*, which has published over many years very useful articles on (as they phrase it) “non-English personal names of a variety of national origins”.

References

1. Black, Bill. 2003. Indexing the Names of Authors from Spanish- and Portuguese-Speaking Countries.

- Science Editor 26(4):118-121.
2. Han Sunghee. 2005. Formats of Korean Authors' Names. Science Editor 28(6):189-190.
 3. International Organization for Standardization (ISO). 2002. Codes for the representation of names of languages — Part 1: Alpha-2 code. ISO 639-1:2002. Geneva, Switzerland: ISO.
 4. International Organization for Standardization (ISO). 2004. Information and documentation — Codes for the representation of names of scripts. ISO 15924:2004. Geneva, Switzerland: ISO. <http://www.unicode.org/iso15924/standard/>.
 5. International Organization for Standardization (ISO). 2009. Linguistic resources management — Multilingual information framework. ISO/DIS 24616 [Draft International Standard]. Geneva, Switzerland: ISO. [PDF]
 6. Internet Assigned Numbers Authority (IANA). IANA Language Subtag Registry. <http://www.iana.org/assignments/language-subtag-registry>.
 7. Internet Engineering Task Force (IETF). 2009. Tags for Identifying Languages, ed. A. Phillips and M. Davis. IETF Best Current Practice (BCP) 47, Network Working Group Request for Comments (RFC) 5646. <http://www.ietf.org/rfc/rfc5646.txt>.
 8. JATS Archiving and Interchange Tag Library, NISO JATS version 0.4. 2011. <http://jats.nlm.nih.gov/archiving/tag-library/0.4/index.html>. March.
 9. JATS Journal Publishing Tag Library, NISO JATS version 0.4. 2011. <http://jats.nlm.nih.gov/publishing/tag-library/0.4/index.html>. March.
 10. JATS Version 3.1 Change Decisions. 2010. NISO Standardized Markup for Journal Articles Working Group Reference Document. August 27.
 11. Kidambi, Misha. 2008. Indian Names: A Guide for Science Editors. Science Editor 31(4):120-120.
 12. Lapeyre, Deborah Aleyne. 2010. Journal Article Tag Sets Version 3.1 Final Modification Decisions. NISO Standardized Markup for Journal Articles Working Group Working Document. July.
 13. Notzon, Beth and Gayle Nesom. 2005. The Arabic Naming System. Science Editor 28(1): 20-21.
 14. Scholarly Publishing in Japanese (SPJ) Working Group. Discussion on defining multilingual data in the NLM DTD, 5 April 2010.
 15. Scholarly Publishing in Japanese (SPJ) Working Group. 2010. Journal Article Tag Sets Version 3.0 Internationalization Modification Requests from the SPJ Working Group (rev 1.0). NISO Standardized Markup for Journal Articles Working Group Working Document. July 22.
 16. Sun Xiao-Ling, and Zhou Jing. 2002. English Versions of Chinese Author's Names in Biomedical Journals: Observations and Recommendations [Dialogue]. Science Editor 25(1):3-4.
 17. The Unicode Standard: Version 6.0 - Core Specification. 2011. The Unicode Consortium. <http://www.unicode.org/versions/Unicode6.0.0/>.
 18. Wikipedia contributors. Burmese name [*sic*]. Wikipedia, The Free Encyclopedia. Accessed July, 2010. http://en.wikipedia.org/wiki/Burmese_names.
 19. Wikipedia contributors. Indian Name [*sic*]. Wikipedia, The Free Encyclopedia. Accessed July, 2010. http://en.wikipedia.org/wiki/Indian_names.
 20. World Wide Web Consortium (W3C). 2008. Extensible Markup Language (XML) 1.0 (Fifth Edition), ed. Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, and François Yergeau. W3C Recommendation 26 November 2008. [Cambridge, Sophia-Antipolis, Tokyo:]: W3C. <http://www.w3.org/TR/2008/REC-xml-20081126/>.
 21. World Wide Web Consortium (W3C). 2001. Ruby Annotation, ed. Marcin Sawicki, Michel Suignard, Masayasu Ishikawa, Martin Dürst, and Tex Texin. W3C Recommendation 31 May 2001. [Cambridge, Sophia-Antipolis, Tokyo:]: W3C. <http://www.w3.org/TR/ruby/>. [Markup errors corrected 25June2008]