XLIFF Extensibility and Metadata

Applied to DITA-XLIFF and Drupal-XLIFF programs

Bryan Schnabel
Content Management Architect, Tektronix,
Chair, OASIS XLIFF TC

@bryanschnabel
I’d like to know how well you know

1. XLIFF
2. DITA
3. Drupal (Web CMS in general)
4. ITS (my money is on this one)
Outline

• Explain XLIFF 1.2 metadata & extensibility method
  – Show this applied to my DITA-XLIFF roundtrip plugin for the Open Toolkit
  – Show this applied to my part of the Drupal-XLIFF module

• Explain XLIFF 2.0 metadata & extensibility method(s)

• Look at this from ITS’s point of view (audience interaction encouraged)
My opinion

Extensibility in XLIFF is not evil
2.5. Extensibility

At times, it may be useful to extend the set of information available in an XLIFF document by inserting constructs defined in various other XML vocabularies. You can add non-XLIFF elements, as well as attributes and attribute values. Adding elements and attributes use the namespace mechanism [XML Names]. Adding attribute values generally involves preceding the value by an "x-" (e.g. <context context-type='x-for-engineers'>).

Although XLIFF offers this extensibility mechanism, in order to avoid a nimiety of information and increase interoperability between tools, it is strongly recommended to use XLIFF capabilities whenever possible, rather than to create non-standard user-defined elements or attributes.

http://docs.oasis-open.org/xliff/v1.2/os/xliff-core.html#Struct_Extension
XLIFF 1.2 is extensible 3 ways

1. Elements: custom namespace, where allowed
2. Attributes: custom namespace, where allowed
3. Attribute values (for XLIFF attributes): concatenate “x-” with attribute string, where allowed
Elements: custom namespace

1. Declare XML namespace:

```xml
<xliff xmlns="urn:oasis:names:tc:xliff:document:1.2"
   xmlns:tek="http://www.tektronix.com"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="urn:oasis:names:tc:xliff:document:1.2
dliff-core-1.2-strict.xsd http://www.tektronix.com
tek_code_trial.xsd" version="1.2">

1. Add non-XLIFF element, where allowed:

```xml
<tek:header>The First Volume of Software Structures</tek:header>
```

http://docs.oasis-open.org/xliff/v1.2/cs02/Sample_AlmostEverything_1.2_strict.xlf
Extensibility is allowed on these XLIFF elements

- xliFF
- header
- tool
- group
- trans-unit
- alt-trans
- bin-unit

(sadly, not on skl or internal-file)

http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html
1. Declare XML namespace:

```xml
<xliff xmlns="urn:oasis:names:tc:xliff:document:1.2"
      xmlns:xmrk="http://www.xmarker.com"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

1. Add non-XLIFF attribute, where allowed:

```xml
<group resname="pubdate" xmrk:ancs="2">
    <trans-unit resname="pubdate" id="pubdate-x-N65545">May 2009</source>
</trans-unit>
</group>
```
Non-XLIFF attributes are allowed on these XLIFF elements

- xlfiff
- file
- tool
- group
- trans-unit
- source
- target
- alt-trans
- bin-unit
- bin-source
- bin-target
- seg-source
- g
- x
- bx
- ex
- ph
- bpt
- ept
- it
- mrk

http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html
Attributes: extensible values

1. For XLIFF attributes with pre-defined values that allow extensible values, concatenate an “x-” prefix to the user-defined value
2. Add the concatenated value to the XLIFF attribute, where allowed

<mrk mtype="x-test">text</mrk>

http://docs.oasis-open.org/xliff/v1.2/cs02/Sample_AlmostEverything_1.2_strict.xlf
Extensible values are allowed in these XLIFF attributes

- alttranstype
- context-type
- count-type
- ctype
- datatype
- purpose
- reformat
- size-unit
- state
- state-qualifier
- unit

http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html
XLIFF 1.2 Extensibility examples

• DITA-XLIFF Roundtrip Open Toolkit plugin
  http://sourceforge.net/projects/ditaxliff/files/
• Drupal XLIFF Tools Module
  http://drupal.org/project/xliff
XLIFF and DITA

(skip tutorial slides on DITA)
DITA/XLIFF Roundtrip

Open Source DITA OT Plugin Demonstration
Choosing the right version of the XLIFF tool

Use my DITA OT Plugin
http://sourceforge.net/projects/ditaxliff/files/

Do not use my other tool, the document-centric xliffRoundTrip Tool
http://sourceforge.net/projects/xliffroundtrip/
A quick sketch of the DITA OT

Skip ahead to SVG
Begin with a set of DITA Topics and Map files
DITA-XLIFF DITA OT Step 1

Tell the DITA OT to create an XLIFF file by pointing at the root map file.
Result of Step 1

You will receive an XLIFF file with all topics & maps, plus a PDF file
DITA-XLIFF DITA OT Step 2

Translate the XLIFF file into the target language

At the transition from the head to the neck there is a recess, a so-called cutaway, at the lower surface, thus comfortable playing is possible in high situations.

Telecaster shares the form of the conventional (acoustic) guitar.

Der Korpus der Telecaster ist der Form
der herkömmlichen (akustischen) Gitarren
nachempfunden. Am Übergang vom Korpus
zum Hals gibt es an der Unterseite eine Aussparung,
das so genannte Cutaway, wodurch ein komfortables
Spiel auch in den hohen Lagen ermöglicht
wird.
DITA-XLIFF DITA OT Step 3

Tell the DITA OT to transform the translated XLIFF into a DITA project.
Result of Step 3

You will receive a translated DITA project (maps, topics, hierarchy)

Skip ahead to SVG
What about the metadata?

• DITA projects are made up of many topic files and map files
• Storing the structure in the skeleton/internal-file element would be ideal
• Custom namespace is not allowed in either of those elements, so . . .
1. Set up the custom namespace in the XSL file

```xml
<xsl:transform
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:opentopic-index="http://www.idiominc.com/opentopic/index"
    xmlns:opentopic="http://www.idiominc.com/opentopic"
    xmlns:ot-placeholder="http://suite-sol.com/namespaces/ot-placeholder"
    xmlns:xmrk="urn:xmarker"
    version="1.0">
```
2. Set up a mode for the skeleton and a different mode for the body

```xml
<xsl:template match="node()|@*" mode="skel">  
  <xsl:copy>  
    <xsl:apply-templates select="@*|node()" mode="skel" />  
  </xsl:copy>  
</xsl:template>

<xsl:template match="node()|@*" mode="body2">  
  <xsl:copy>  
    <xsl:apply-templates select="@*|node()" mode="body2" />  
  </xsl:copy>  
</xsl:template>
```

3. Apply format templates to skel mode, and transform templates to body mode

```xml
<xsl:apply-templates mode="skel" />  
<xsl:apply-templates mode="body2" />  
```
Result: all format metadata in custom namespace
Result: all translation units in XLIFF namespace
XLIFF and Web Content Management Systems
XLIFF in Drupal

About Us, in this case, means “about me”

I work for Tektronix as an Information Architect. I also chair the OASIS XLIFF Technical Committee. For 10 years I was a head coach for various youth hockey teams. Most recently I was the head coach for the Sherwood High School hockey team (5 seasons). I have not coached for a few years now, but I have a feeling I might coach again in the coming years. I am a ski enthusiast (season ticket holder each year at Mt. Hood Meadows in Oregon). My latest hobby (obsession) is to try to write and record an album of all these songs rattling around in my head (write the songs, play all the instruments, record using Logic software). I seem to be collecting guitars lately. My goals is the actually release a CD.

I have a wife and two children. My wife and I live in Sherwood Oregon. My son lives in Eugene Oregon (just graduated from the University of Oregon, preparing for Law School there). My daughter lives in Portland Oregon and attends Portland State University (pre-med).

Finding out even more about me
It’s funny, I have three social media accounts, and on each, I have a distinct “persona” I’m trying to pull off.

My LinkedIn Connection
You can find me at Bryan Schnabel on linkedin. I try my best to be a respectable business type. Although I do use a picture of me playing bass. I head the very serious XLIFF group on linkedin. But on the other hand, I also head the “Respectable Business People Who Love Raucoos Music” group.

My MySpace Music Connection
You can find me at bryanschnabelmusik. I post my songs, try to act cool, and do the usual bio stuff. In some cases I think it would be dreadful if my linkedin crowd ever met my myspace music crowd.
Drupal XLIFF declares custom namespace

<xliiff xmlns="urn:oasis:names:tc:xliiff:document:1.2"
xmlns:xmrk="urn:xmarker"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:oasis:names:tc:xliiff:document:1.2
xliiff-core-1.2-strict.xsd
urn:xmarker
xmarker.xsd"
version="1.2">
Result: all format metadata in custom namespace

```xml
<file datatype="plaintext" source-language="en" original="html">
    <header>
        <xmrk:nest>
            <xmrk:html xmarker_idref="html-0" node="795256-795586">
                <xmrk:title xmarker_idref="title-1"/>
                <xmrk:body xmarker_idref="body-2"/>
                <xmrk:teaser xmarker_idref="teaser-3"/>
                <xmrk:field_part_number xmarker_idref="field_part_number-4"/>
                <xmrk:field_product_number xmarker_idref="field_product_number-5"/>
                <xmrk:field_metatag_code xmarker_idref="field_metatag_code-6"/>
                <xmrk:field_import_name xmarker_idref="field_import_name-7"/>
                <xmrk:field_unique_name xmarker_idref="field_unique_name-8"/>
                <xmrk:field_datasheet_specifications xmarker_idref="field_datasheet_specifications-9"/>
                <xmrk:field_omar_inventory_status xmarker_idref="field_omar_inventory_status-10"/>
                <xmrk:field_omar_mktg_status xmarker_idref="field_omar_mktg_status-11"/>
                <xmrk:field_record_length xmarker_idref="field_record_length-12"/>
                <xmrk:field_analog_channel xmarker_idref="field_analog_channel-13"/>
                <xmrk:field_digital_channel xmarker_idref="field_digital_channel-14"/>
                <xmrk:field_analog_bandwidth xmarker_idref="field_analog_bandwidth-15"/>
                <xmrk:field_digital_analysis_tools xmarker_idref="field_digital_analysis_tools-16"/>
                <xmrk:field_max_waveform_capture_rate xmarker_idref="field_max_waveform_capture_rate-17"/>
                <xmrk:field_optional_analysis xmarker_idref="field_optional_analysis-18"/>
                <xmrk:field_rf_channel xmarker_idref="field_rf_channel-19"/>
                <xmrk:field_rf_frequency xmarker_idref="field_rf_frequency-20"/>
                <xmrk:field_trigger_types xmarker_idref="field_trigger_types-21"/>
                <xmrk:field_waveform_maths_analysis xmarker_idref="field_waveform_maths_analysis-22"/>
                <xmrk:field_sample_rate xmarker_idref="field_sample_rate-23"/>
                <xmrk:field_rise_time xmarker_idref="field_rise_time-24"/>
                <xmrk:field_other_features xmarker_idref="field_other_features-25"/>
                <xmrk:field_connectivity xmarker_idref="field_connectivity-26"/>
                <xmrk:field_warranty xmarker_idref="field_warranty-27"/>
            </xmrk:html>
        </xmrk:nest>
    </header>
</file>
```
Result: all translation units in XLIFF namespace

Note: not all metadata needed to go into a custom namespace
Drupal creates XLIFF

<group id="id2546682axmarkhtml-0">
<group id="id2546683bxmarkhead-1">
<trans-unit id="title-2">
<source>Drupal creates XLIFF</source>
<target>Drupal erstellt XLIFF</target>
</trans-unit>
</group>
</group>

<group id="id2546686bxmarkbody-3">
<trans-unit id="h2-4">
<source>Drupal creates XLIFF</source>
<target>Drupal erstellt XLIFF</target>
</trans-unit>
<br/>
<trans-unit id="p-5">
<source>I work for Tektronix as an Information Architect. I also chair the<br/>
<target>Ich arbeite als IT-Architekt für Tektronix. Ich bin außerdem Vorsitzender</target>
</trans-unit>
<br/>
<trans-unit id="p-9">
<source>I have a wife and two children. My wife and I live in Sherwood Oregon. My son lives in Eugen<br/>
<target>Ich bin verheiratet und habe zwei Kinder. Meine Frau und ich leben in Sherwood im US-Bundess</target>
</trans-unit>
<br/>
<trans-unit id="h3-12">
<source>Drupal creates XLIFF</source>
<target>Drupal erstellt XLIFF</target>
</trans-unit>
<br/>
<trans-unit id="p-13">
<source>Its funny, I have three social media accounts, and on each, I have a distinct "persona". I'm<br/>
<target>Lustigerweise habe ich Konten in drei verschiedenen sozialen Netzwerken und trete in jedem d</target>
</trans-unit>
<br/>
<trans-unit id="h3-14">
<source>Drupal creates XLIFF</source>
<target>Drupal erstellt XLIFF</target>
</trans-unit>
<br/>
</group>
Then import the translated XLIFF
How should XLIFF 2.0 support metadata & extensibility

- After many months of thoughtful and spirited debate, a vote was taken:
  - Custom namespaces (XLIFF 1.2 method)
  - In XLIFF metadata elements (example on next slide)?
  - In custom namespaces and XLIFF elements?
Proposed XLIFF elements for extensibility

• For this markup

<para id="g_3423_spectrum" alt="It's orders of magnitude faster" rev="c">This is orders of magnitude faster than swept analysis techniques.</para>

https://wiki.oasis-open.org/xliff/XLIFF2.0/Feature/Preserve%20XML%20attribute%20or%20metadata%20without%20extensibility
Proposed XLIFF elements for extensibility

• Elements *something like this*

```xml
<unit id="a1">
  <segment>
    <source> This is orders of magnitude faster than swept analysis techniques. </source>
  </segment>
  <metaHolder type="x-attributes">
    <meta key="rev">c</meta>
    <meta key="id">g_3423_spectrum</meta>
    <meta key="alt" id="a1.a">It's orders of magnitude faster</meta>
  </metaHolder>
  <metaHolder type="count">
    <meta key="words">2</meta>
    <meta key="chars">13</meta>
  </metaHolder>
</unit>
<unit id="a2" idref="a1.a">
  <segment>
    <source> It's orders of magnitude faster </source>
  </segment>
</unit>
```

Note: while this is the sample of record in the XLIFF wiki, improvements/refinements are on record on the mailing list

https://wiki.oasis-open.org/xliff/XLIFF2.0/Feature/Preserve%20XML%20attribute%20or%20metadata%20without%20extensibility
The results?

- Two-way tie
  - Elements only
  - Elements and custom namespaces
- More debate followed, and a tie-breaker ballot is started, results pending
You can have a voice

Engagement from the community

• Join the TC (https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff)
• Or follow the TC mailing list (https://lists.oasis-open.org/archives/xliff/)
• Follow the TC wiki (https://wiki.oasis-open.org/xliff/)
• You can comment, even in you’re not on the TC* (https://lists.oasis-open.org/archives/xliff-comment/)
• PLEASE do not just stand on the side lines grousing

* We do monitor, care about, and respond to the comment list
Thank you

Bryan Schnabel
bschnabel@bschnabel.com