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

XLIFF 1.2 Extensibility

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:

```
<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 xliff-core-1.2-strict.xsd http://www.tektronix.com
    tek_code_trial.xsd" version="1.2">
```

1. Add non-XLIFF element, where allowed:

```
<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

Attributes: custom namespace

1. Declare XML namespace:

```
<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"
    xsi:schemaLocation="urn:oasis:names:tc:xliff:document:
    1.2 xliff-core-1.2-strict.xsd http://www.xmarker.com
    xmrk.xsd" version="1.2">
```

1. Add non-XLIFF attribute, where allowed:

```
<group resname="pubdate" xmrk:ancs="2">
 <trans-unit resname="pubdate" id="pubdate-x-N65545“>
  <source xmrk:ancs="2">May 2009</source>
 </trans-unit>
</group>
```

Non-XLIFF attributes are allowed on these XLIFF elements

- xliff
- 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

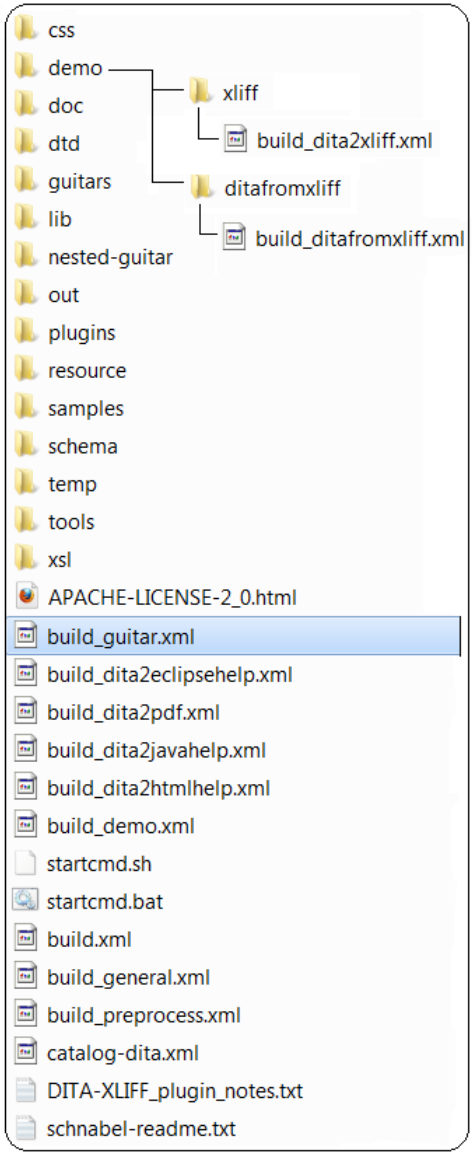
Use my *DITA OT Plugin*

http://sourceforge.net/projects/ditaxliff/files/

Do not use my other tool, the document-centric *xliffRoundTrip Tool*

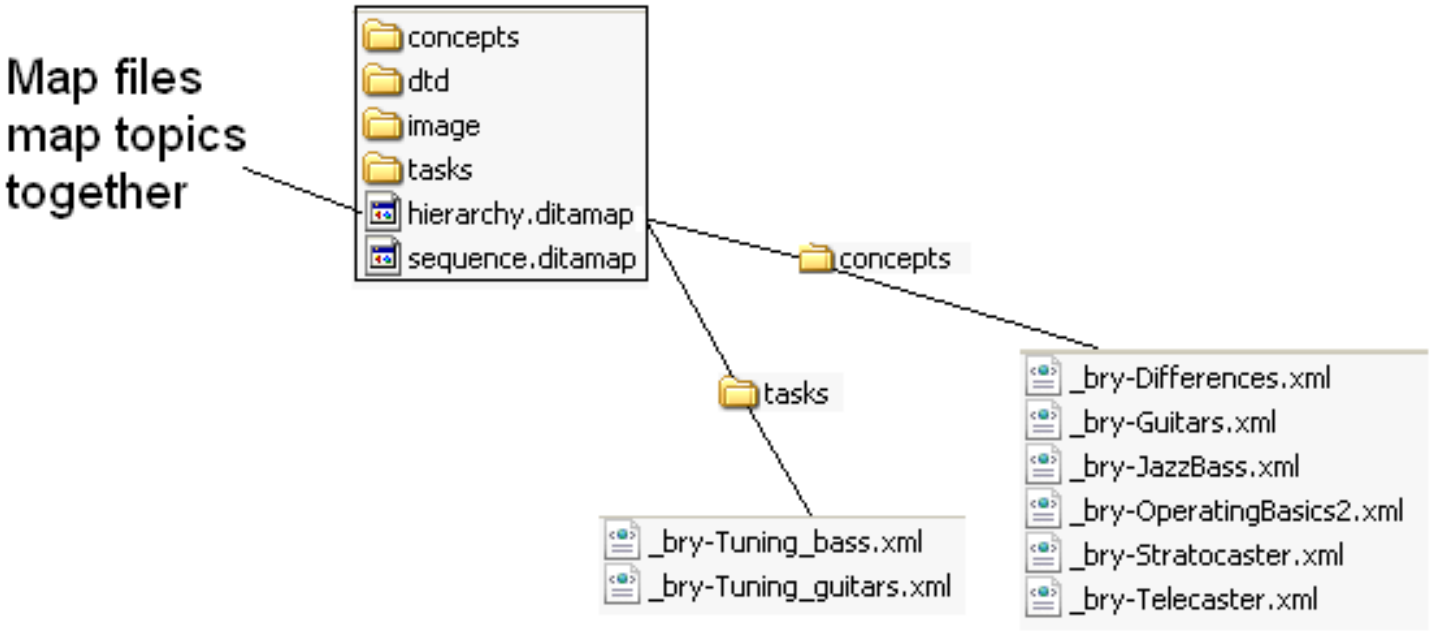
http://sourceforge.net/projects/xliffroundtrip/

A quick sketch of the DITA OT



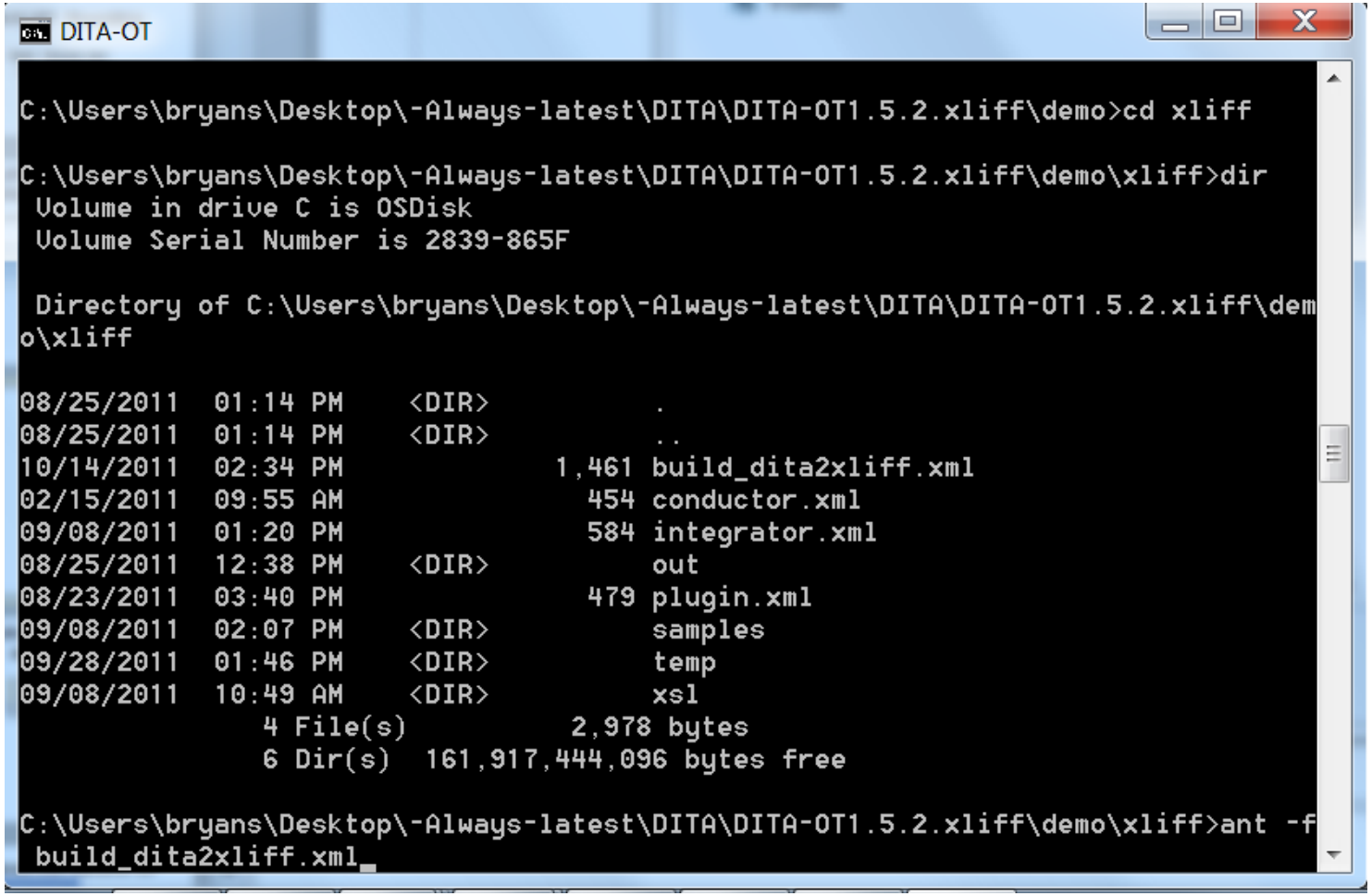
Skip ahead to SVG

Begin with a set of DITA Topics and Map files



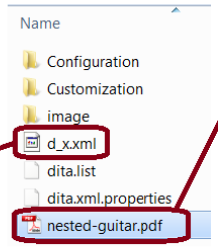
DITA-XLIFF DITA OT Step 1

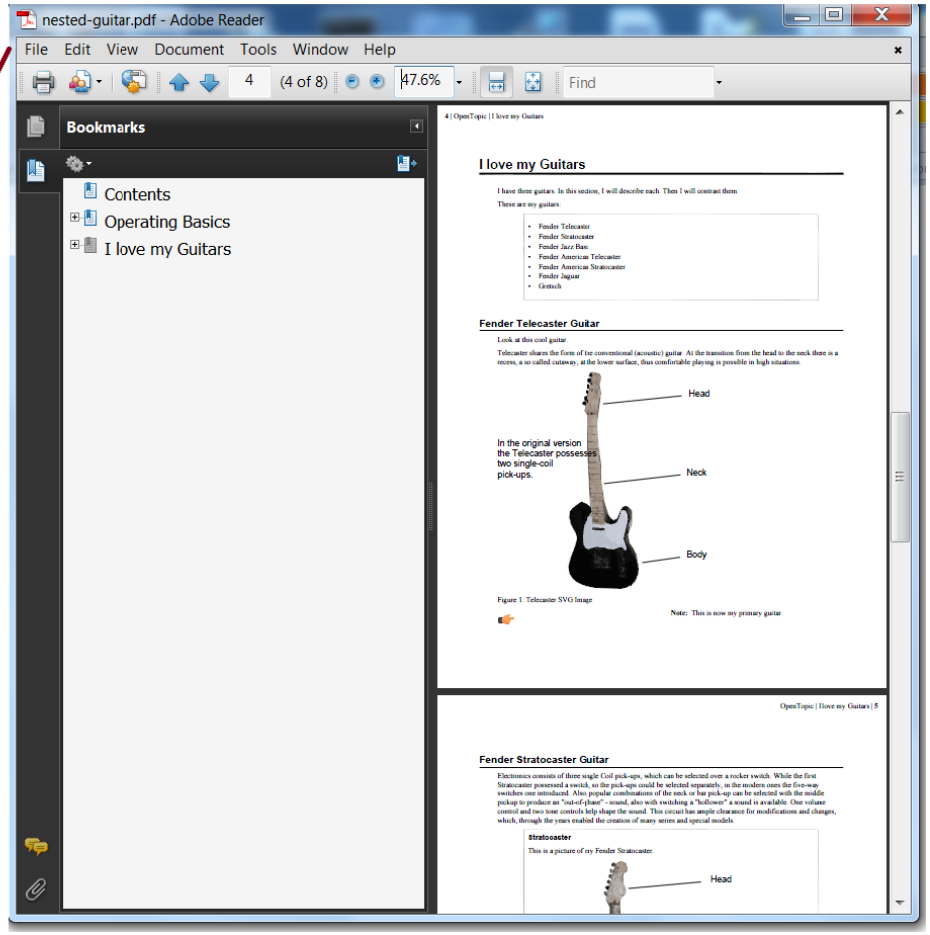
Tell the DITA OT to create an XLIFF file by pointing at the root map file



Result of Step 1

You will receive an XLIFF file with all topics & maps, plus a PDF file






```
<body><!-- (c) Bryan Schnabel 2008 --><group id="d10e2bxmarkconcept-0">
      <trans-unit id="title-1">
         <source>Operating Basics</source>
         <target>Operating Basics</target>
      </trans-unit>
      <group id="d10e5bxmarkconbody-2">
         <trans-unit id="p-3">
            <source>Guitars are played by strumming or plucking the strings.</source>
            <target>Guitars are played by strumming or plucking the strings.</target>
         </trans-unit>
      </group>
      <group id="d10e8bxmarkconcept-4">
         <trans-unit id="title-5">
            <source>Tuning Guitars</source>
            <target>Tuning Guitars</target>
         </trans-unit>
         <group id="d10e11bxmarkconbody-6">
            <trans-unit id="p-7">
               <source>Electric guitars are tuned in many ways. The most common is standard
urce>
               <target>Electric guitars are tuned in many ways. The most common is standard
rget>
            </trans-unit>
            <trans-unit id="p-8">
               <source>Electric bass guitars are usually tuned "EADG."</source>
               <target>Electric bass guitars are usually tuned "EADG."</target>
            </trans-unit>
         </group>
      </group>
   </group>
</body>
```

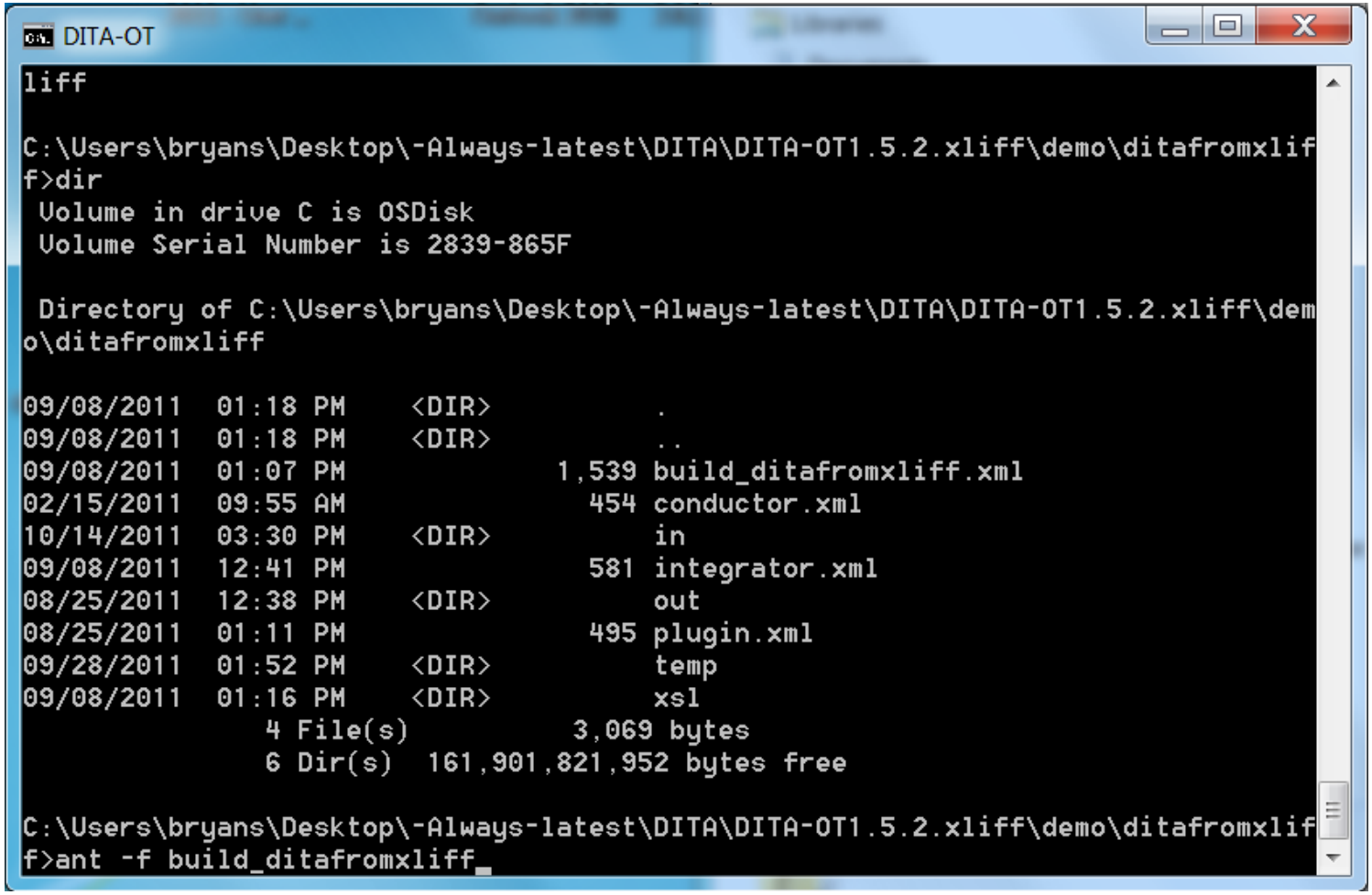
DITA-XLIFF DITA OT Step 2

Translate the XLIFF file into the target language

```
<file datatype="xml" original="concepts/_bry-Telecaster.dita" xmrk:level="map" xmrk:order="2" source-language="en-us" target-langu
   <body>
      <group resname="concept" xmrk:rootname="concept" xmrk:id="Fender_Telecaster_Guitar" xml:lang="en-us"/>
      <group resname="topicref" xmrk:ancs="1" xmrk:one="1" xmrk:href="concepts/_bry-Telecaster.dita" xmrk:type="concept">
         <trans-unit resname="title" id="title-x-mch1-d3e3" xmrk:ancs="1">
            <source xmrk:ancs="1" xml:lang="en-us">Fender Telecaster Guitar</source>
            <target xml:lang="de">Fender Telecaster Guitar</target>
         </trans-unit>
         <group resname="conbody" xmrk:ancs="1">
            <group id="throw1-d3e5" xmrk:ancs="1" resname="conbody">
               <trans-unit resname="p" id="p-x-mch1-d3e6" xmrk:ancs="2">
                  <source xmrk:ancs="2" xml:lang="en-us">Look at this cool guitar.</source>
                  <target xml:lang="de">Look at this cool guitar.</target>
               </trans-unit>
               <trans-unit resname="p" id="p-x-mch1-d3e8" xmrk:ancs="2">
                  <source xmrk:ancs="2" xml:lang="en-us">Telecaster shares the form of the conventional (acoustic) guitar.
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.</source>
                  <target xml:lang="de">Der Korpus der Telecaster ist der Form
der herkÃ¶mmlichen (akustischen) Gitarre
nachempfunden.  Am Ãbergang vom Korpus
zum Hals gibt es an der Unterseite eine Aussparung,
das so genannte Cutaway, wodurch ein komfortables
Spiel auch in den hohen Lagen ermÃ¶glicht
wird.</target>
               </trans-unit>
               <group resname="fig" xmrk:ancs="2">
                  <group id="throw1-d3e10" xmrk:ancs="2" resname="fig">
                     <group id="sl11p2d3e11" xmrk:ancs="3">
                        <trans-unit id="d3e11" xmrk:ancs="3">
                           <source xml:lang="en-us">
                              <x id="image-x-mch2-d3e11" xmrk:ancs="3" xmrk:href="../image/tele.xml.svg"/>
                           </source>
                           <target xml:lang="de">
                              <x id="image-x-mch2-d3e11" xmrk:ancs="3" xmrk:href="../image/tele.xml.svg"/>
                           </target>
                        </trans-unit>
                     </group>
                     <trans-unit resname="p" id="p-x-mch1-d3e12" xmrk:ancs="3">
                        <source xmrk:ancs="3" xml:lang="en-us">Figure 1: Telecaster SUG Image</source>
                        <target xml:lang="de">Figure 1: Telecaster SUG Image</target>
                     </trans-unit>
                  </group>
               </group>
               <trans-unit resname="note" id="note-x-mch1-d3e14" xmrk:ancs="2">
```

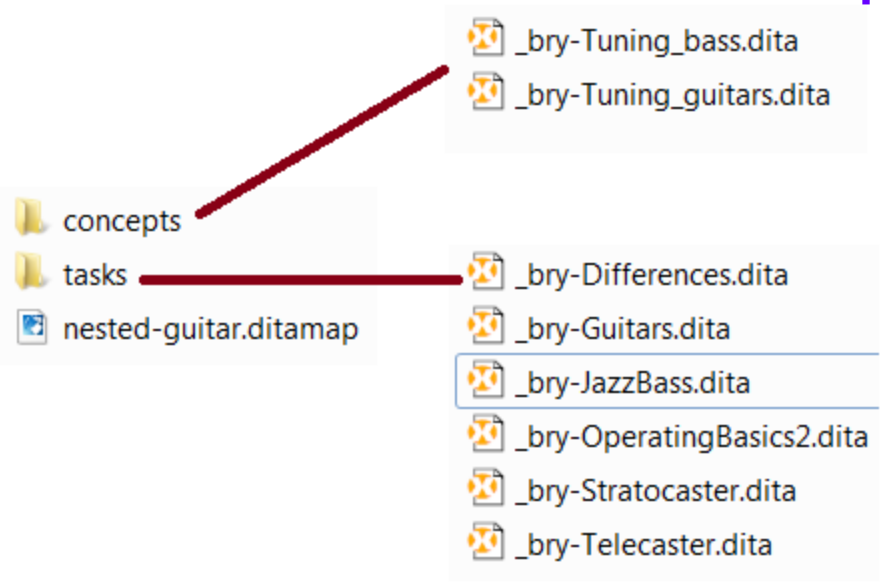
DITA-XLIFF DITA OT Step 3

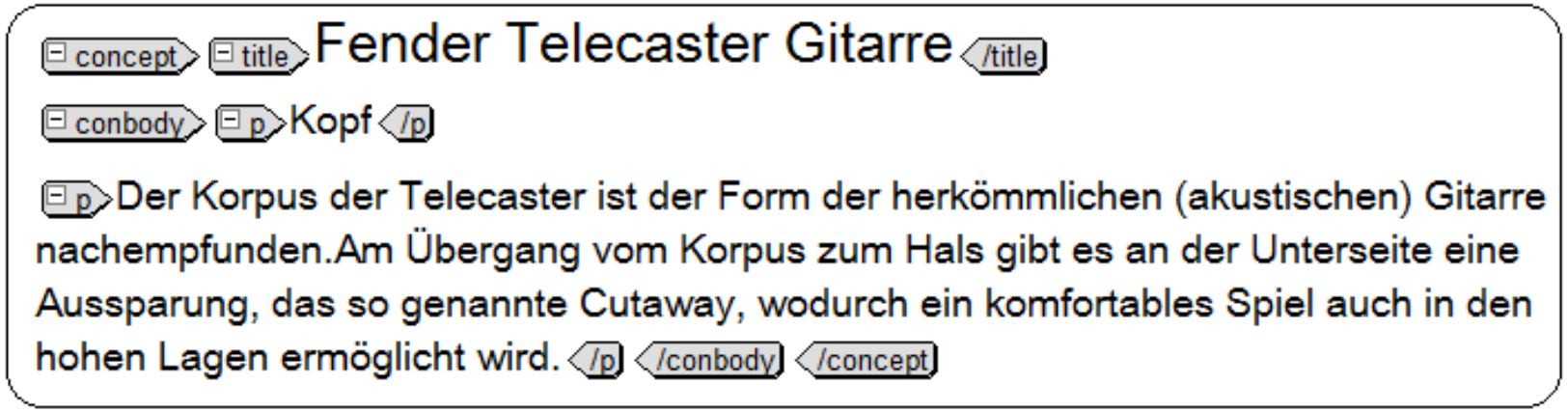
Tell the DITA OT to transform the translated XLIFF into a DITA project



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 . . .

DITA-OT1.5.4.xliff\demo\xliff\xsl

1. Set up the custom namespace in the XSL file

```
<xsl:transform
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:opentopic-index="http://www.idiominc.com/opentopic/index"
  xmlns:ditaarch="http://dita.oasis-open.org/architecture/2005/"
  xmlns:opentopic="http://www.idiominc.com/opentopic"
  xmlns:ot-placeholder="http://suite-sol.com/namespaces/ot-
placeholder"
  xmlns:xmrk="urn:xmarker"
  version="1.0">
```

DITA-OT1.5.4.xliff\demo\xliff\xsl

2. Set up a mode for the skeleton and a different mode for the body

```
<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

```
<xsl:apply-templates mode="skel" />
<xsl:apply-templates mode=“body2" />
```

Result: all format metadata in custom namespace

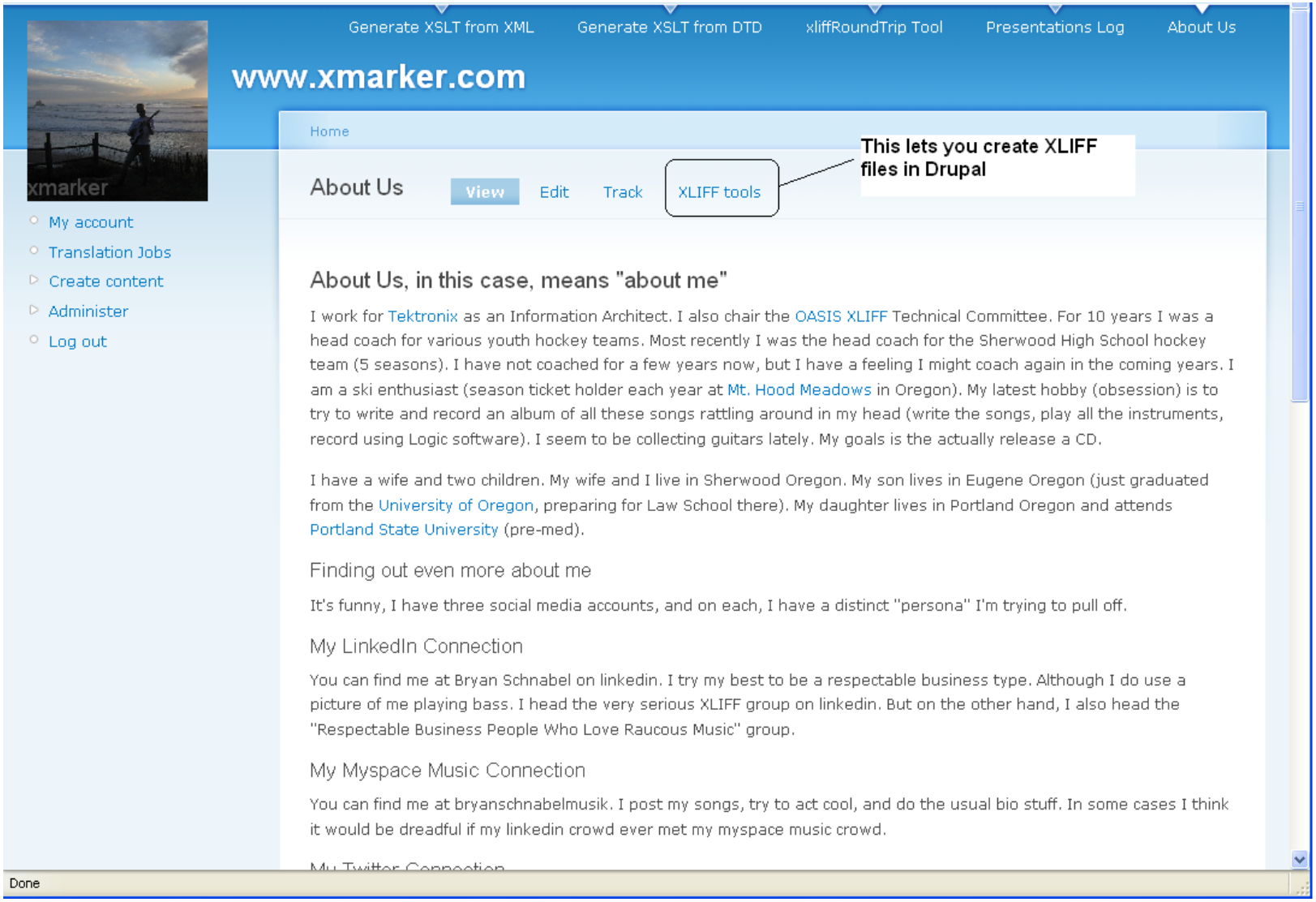
```
<file datatype="plaintext" source-language="en"
      original="C:\Users\bryans\Desktop\-Always-latest\DITA\DITA-OT1.5.4.xliff\demo\xliff\samples\ComstartUserGuide\ComstarUserGuide.ditamap"
      build-num="d1e1">
   <header>
      <xmrk:nest>
         <xmrk:type u_id="Comstaruserguide">map</xmrk:type>
         <xmrk:outputURI>ComstarUserGuide.ditamap</xmrk:outputURI>
         <xmrk:file xtrf="C:\Users\bryans\Desktop\-Always-latest\DITA\DITA-OT1.5.4.xliff\demo\xliff\samples\ComstartUserGuide\ComstarUserGuide.ditamap">
            <xmrk:map xmarker_idref="map-0" id="Comstaruserguide" ditaarch:DITAArchVersion="1.2"
                      domains="(topic delay-d)                    (map mapgroup-d)                    (topic indexing-d)
                      class="- map/map ">
               <xmrk:title xmarker_idref="title-1" class="- topic/title "/>
               <xmrk:topicmeta xmarker_idref="topicmeta-2" class="- map/topicmeta ">
                  <xmrk:copyright xmarker_idref="copyright-3" class="- topic/copyright ">
                     <xmrk:copyryear xmarker_idref="copyryear-4" year="2011" class="- topic/copyryear "/>
                     <xmrk:copyrholder xmarker_idref="copyrholder-5" class="- topic/copyrholder "/>
                  </xmrk:copyright>
               </xmrk:topicmeta>
               <xmrk:keydef xmarker_idref="keydef-6" keys="glossary" href="Glossgroup.dita"
                            processing-role="resource-only"
                            class="+ map/topicref mapgroup-d/keydef "/>
               <xmrk:topicref xmarker_idref="topicref-7" href="ComstarPreface.dita" toc="no"
                              class="- map/topicref ">
                  <xmrk:topicref xmarker_idref="topicref-8" href="WarningsCode.dita" toc="no"
                                 class="- map/topicref "/>
               </xmrk:topicref>
               <xmrk:topicref xmarker_idref="topicref-9" href="ComstarOverview.dita" format="dita"
                              scope="local"
                              toc="no"
                              class="- map/topicref ">
                  <xmrk:topicref xmarker_idref="topicref-10" href="ComstarOverviewCode.dita" format="dita"
                                 scope="local"
                                 toc="no"
                                 class="- map/topicref "/>
               </xmrk:topicref>
               <xmrk:topicref xmarker_idref="topicref-11" href="QuickGuide.dita" navtitle="Quick start guide"
                              format="dita"
                              scope="local"
                              type="reference"
```

Result: all translation units in XLIFF namespace

```
227     <body>
228       <group id="d6e1bxmarkmap-0">
229         <trans-unit id="title-1">
230           <source>Comstar User Guide</source>
231           <target>Comstar User Guide</target>
232         </trans-unit>
233         <group id="d6e4bxmarktopicmeta-2">
234           <group id="d6e5bxmarkcopyright-3">
235             <group id="d6e6cxmarkcopyryear-4"/>
236             <trans-unit id="copyrholder-5">
237               <source>Comstar Phones</source>
238               <target>Comstar Phones</target>
239             </trans-unit>
240           </group>
241         </group>
242         <group id="d6e9cxmarkkeydef-6"/>
243         <group id="d6e10bxmarktopicref-7">
244           <group id="d6e11cxmarktopicref-8"/>
245         </group>
246         <group id="d6e12bxmarktopicref-9">
247           <group id="d6e13cxmarktopicref-10"/>
248         </group>
249         <group id="d6e14bxmarktopicref-11">
250           <group id="d6e15bxmarktopicmeta-12">
251             <trans-unit id="navtitle-13">
252               <source>Get Started Now</source>
253               <target>Get Started Now</target>
254             </trans-unit>
255           </group>
```

XLIFF and Web Content Management Systems

XLIFF in Drupal



Drupal XLIFF declares custom namespace

```
<xliff xmlns="urn:oasis:names:tc:xliff:document:1.2"
       xmlns:xmrk="urn:xmarker"
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xsi:schemaLocation="urn:oasis:names:tc:xliff:documen
t:1.2
       xliff-core-1.2-strict.xsd
       urn:xmarker
       xmarker.xsd"
       version="1.2">
```

Result: all format metadata in custom namespace

```
<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_math_analysis xmarker_idref="field_waveform_math_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>
```

Result: all translation units in XLIFF namespace

```
43      <body>
44         <group id="id3655598axmarkhtml-0">
45            <trans-unit id="title-1">
46               <source>P6717</source>
47               <target>P6717</target>
48            </trans-unit>
49            <trans-unit id="body-2">
50               <source>17 channel general purpose logic probe for MSO70000 with accessories</source>
51               <target>17 channel general purpose logic probe for MSO70000 with accessories</target>
52            </trans-unit>
53            <group id="id3656367exmarkteaser-3"/>
54            <group id="id3656369exmarkfield_part_number-4"/>
55            <group id="id3656371exmarkfield_product_number-5"/>
56            <group id="id3656374exmarkfield_metatag_code-6"/>
57            <group id="id3656376exmarkfield_import_name-7"/>
58            <group id="id3656378exmarkfield_unique_name-8"/>
59            <group id="id3656380exmarkfield_datasheet_specifications-9"/>
60            <trans-unit id="field_omar_inventory_status-10">
61               <source>Active</source>
62               <target state="final">Active</target>
63            </trans-unit>
64            <trans-unit id="field_omar_mktg_status-11">
65               <source>LS</source>
66               <target state="final">LS</target>
67            </trans-unit>
```

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

Drupal creates XLIFF

```
<body>
  <group id="id2546682axmarkhtml-0">
    <group id="id2546683bxmarkhead-1">
      <trans-unit id="title-2">
        <source>About Us</source>
        <target>Wer wir sind</target>
      </trans-unit>
    </group>
    <group id="id2546686bxmarkbody-3">
      <trans-unit id="h2-4">
        <source>About Us, in this case, means "about me"</source>
        <target>"Wir", das bedeutet in diesem Fall eigentlich "ich".</target>
      </trans-unit>
      <trans-unit id="p-5">
        <source>I work for <g id="a-6">Tektronix</g> as an Information Architect. I also chair the <g id="a-
        <target>Ich arbeite als IT-Architekt fÃ¼r <g id="a-6">Tektronix</g>. Ich bin auÃ■erdem Vorsitzender
      </trans-unit>
      <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
        <target>Ich bin verheiratet und habe zwei Kinder. Meine Frau und ich leben in Sherwood im US-Bundess
      </trans-unit>
      <trans-unit id="h3-12">
        <source>Finding out even more about me</source>
        <target>Wenn Sie noch mehr Ã¼ber mich wissen mÃ¶chten</target>
      </trans-unit>
      <trans-unit id="p-13">
        <source>It's funny, I have three social media accounts, and on each, I have a distinct "persona" I'm
        <target>Lustigerweise habe ich Konten in drei verschiedenen sozialen Netzwerken und trete in jedem d
      </trans-unit>
      <trans-unit id="h3-14">
        <source>My LinkedIn Connection</source>
        <target>Mein LinkedIn-Kanal</target>
```

Translator reads the ***source*** element

Enters the translation in the ***target*** element

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*

```
<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