

# Standards-based Translation with W3C ITS and OASIS XLIFF



Christian Lieske (SAP AG)  
Felix Sasaki (Fachhochschule Potsdam)  
Yves Savourel (Enlaso)  
Bryan Schnabel (Tektronix)

**tcworld**  
conference **2009**

Rhein-Neckar-Hallen Wiesbaden  
Thursday, 5th November 2009  
8:45 - 10:30 am, Room 1A/3

# Presenters

## Prof. Dr. Felix Sasaki



Univ. of Applied Sciences  
Fac. of Information Science

- Appointed to Prof. in 2009
- Head of the German-Austrian W3C-Office
- Before, staff of the World Wide Web Consortium (W3C) in Japan
- Main field of interest: combined application of W3C technologies for representation and processing of multilingual information
- Studied Japanese, Linguistics and Web technologies at various Universities in Germany and Japan

## Christian Lieske



Globalization Services  
SAP AG

- Knowledge Architect
- Content engineering and process automation (including evaluation, prototyping and piloting)
- Main field of interest: Internationalization, translation approaches and natural language processing
- Contributor to standardization at World Wide Web Consortium (W3C) OASIS and elsewhere
- Degree in Computer Science with focus on Natural Language Processing and Artificial Intelligence

# Contributors

## Yves Savourel



### ENLASO Corporation

- Localization Solutions Architect
- Chaired the Internationalization Tag Set Working Group at the W3C
- Author of the book *XML Internationalization and Localization*
- In the localization industry for more than 15 years; part of several efforts to take advantage of XML in localization
- One of the architects of *XLIFF* and *TMX*

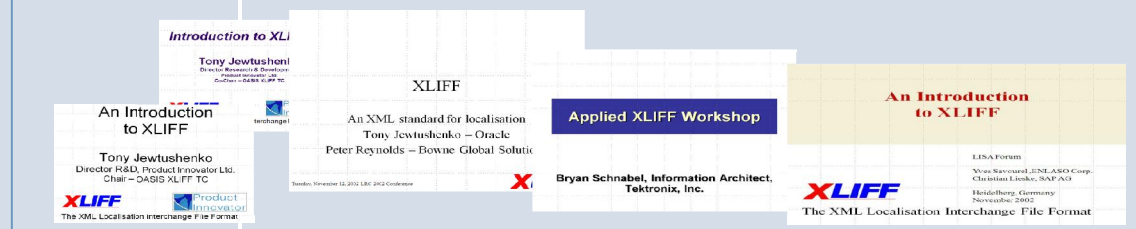
## Bryan Schnabel



### Tektronix

- XML Information Architect
- Chairs the XLIFF Technical Committee at OASIS
- Part of several efforts to take advantage of XML in localization

This presentation draws on the work of other ITS and XLIFF experts



Special thanks: Richard Ishida , Tony Jewtuschenko, Peter Reynolds

# Expectations?!

## You expect ...

Demonstration of a specific format, solution, method or procedure in practice.

An offering related to localization

**See how standards help and interact during translation processes**

A quiet place far away from the fair ;-)

**A tutorial designed for the professional level audience**

**A chance to use XML-related skills/knowledge (markup, XSLT, Xpath)**

Basics of W3C ITS and OASIS XLIFF

**That's what we expected ...**

# Agenda

1. **Challenges with Proprietary Globalization**
2. **Format 1: W3C Internationalization Tag Set (ITS)**
3. **Format 2: OASIS XML Localization Interchange File Format (XLIFF)**
4. **The Relationship between ITS and XLIFF**
5. **Tool: XLIFF-related converters based on ITS**
6. **Q&A**

1. **Challenges with Proprietary Globalization**
2. Format 1: W3C Internationalization Tag Set (ITS)
3. Format 2: OASIS XML Localization Interchange File Format (XLIFF)
4. The Relationship between ITS and XLIFF
5. Tool: XLIFF-related converters based on ITS
6. Q&A

# CHALLENGES WITH PROPRIETARY GLOBALIZATION

Translation is an intrinsic part of anything related to globalization. Internationalization is another one.

The core internationalization and translation tasks, are only abstractions for steps in a series of activities in which many actors participate.

Globalization-related code design and process design are challenging.

# Quiz (1)

Scenario: Configure a spell checker, so that only natural language content is being considered by the checker. Answer a couple of questions for getting the configuration right.

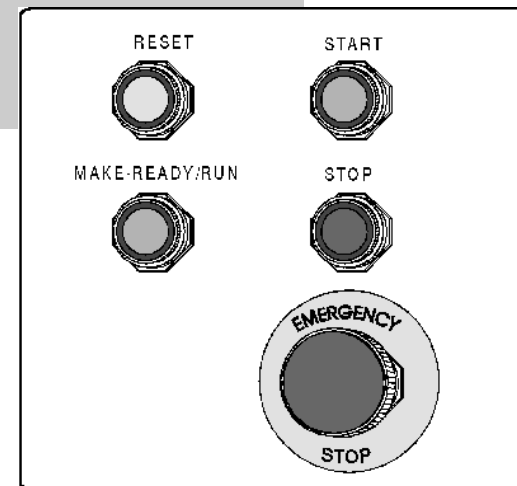
Language of the content?	<Собирание версия="1.2-3">
Terms?	<Объект id="12"> <НомерОбъекта>OnlineCard</НомерОбъекта> <ВНаличии>123</ВНаличии>
Codes?	<Описание xml:lang="ja">第二発電機</Описание> </Объект>
Footnotes?	<Объект id="64"> <НомерОбъекта>45-7894-456</НомерОбъекта> <ВНаличии>Latest Offer</ВНаличии>
Foreign language expressions?	<Оп xml:lang="ja">手動ウォーター・ポンプ</Оп> </Объект>
Annotations for readers?	</Собирание>

# Quiz (2)

Scenario: Translate

Case for translated „uitext“?

```
<para>  
Press the  
<uitext>START</uitext>  
button to sound the horn. The  
<uitext>READY/RUN</uitext>  
indicator flashes.  
</para>
```



Adapted from Richard Ishida (W3C)

# Quiz (3)

## Scenario: Translate

Language of the content?

Terms?

Codes?

Footnotes?

Parts to stay in source language?

Annotations for readers?

```
<resources>  
<section id="Homepage" >  
  <arguments>  
    <string>page</string>  
    <string>childlist</string>  
  </arguments>  
  <variables>  
    <string>POLICY</string>  
    <string>Corporate Policy</string>  
  </variables>  
  <keyvalue_pairs>  
    <string>ABC Corp. - Policy Repository</string>  
    <string>Footer_Last</string>  
    <string>List of Available Policies</string>  
  </keyvalue_pairs>  
</section>  
</resources>
```

## Quiz (4)

Scenario: Decide (for example in your role as source content quality manager) whether the markup is correct.

The title says "W3C ,סוּאִיבָה תּוֹלִיעַפּ, W3C" in Hebrew.

Correct markup?

```
<p>  
The title says "  
<quote xml:lang="he"> ,סוּאִיבָה תּוֹלִיעַפּ, W3C</quote>  
" in Hebrew.  
</p>
```

XHTML:

```
The title says "  
<span dir="rtl"> ,סוּאִיבָה תּוֹלִיעַפּW3C</span>  
" in Hebrew.
```

Adapted from Richard Ishida (W3C) <http://www.w3.org/International/questions/qa-bidi-controls>

## Quiz (5)

Scenario: Decide (for example in your role as source content quality manager) whether the markup is suitable for the overall process chain.

Volcanic eruptions have literally devastated large inhabited areas. During the 1914 eruption of Sakurajima in Kyushu, 687 houses in Kurokami were buried in hot ash. What remained of this shrine gate, previously five meters tall, was left as a reminder.



*Kurokami maibutsu gate  
(腹五社神社黒神埋没鳥居),  
Sakurajima Island.*

Suitable markup?

```
<image src="kk-torii.jpg" height="180"  
width="240" caption="Kurokami maibutsu gate  
(腹五社神社黒神埋没鳥居), Sakurajima Island." />
```

Better:

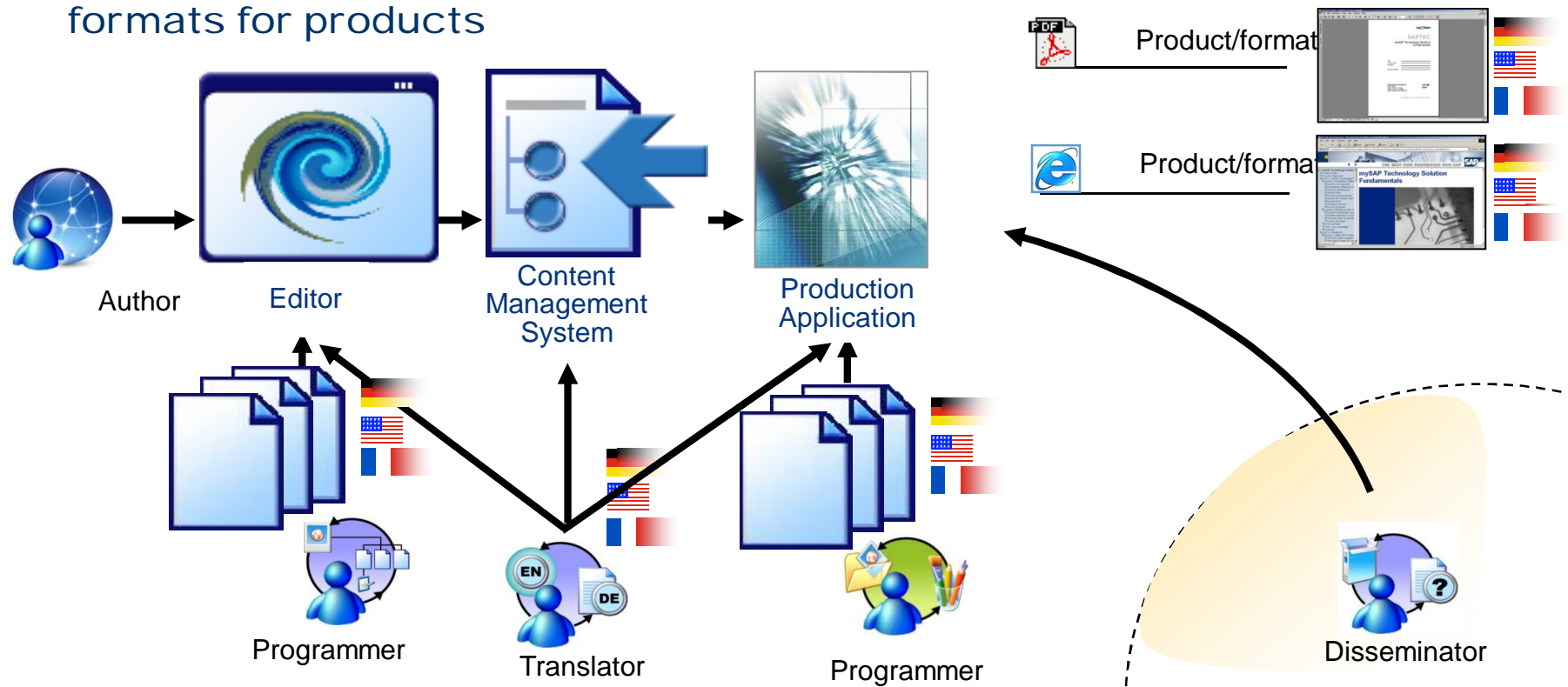
```
<image src="kk-torii.jpg" height="180" width="240">  
<caption>  
Kurokami maibutsu gate (  
<span xml:lang="ja">腹五社神社黒神埋没鳥居</span>  
>), Sakurajima Island.  
</caption>  
</image>
```

Adapted from Richard Ishida (W3C)

# The Real World I

Four areas are substantial for globalization/translation processes:

1. Source content
2. Collaborative work
3. Coupled applications
4. Languages and formats for products



## Content

- Editing format(s) for texts (eg. .doc, .dita)
- Other editing format(s) (eg. for graphics)
- Possible editing format(s) for translation
- Content architecture (eg. relationships between objects, and composition)<sup>1</sup>
- Meta-data (for internationalization and localization as well as for professional service delivery)

## Processes

- For overall professional services delivery translation (PSDT)
- For core translation tasks
- Related to flow of content
- Related to flow of information for context activities such as billing
- Related to production-related aspects such as forecasting and reporting
- Gates and actors for PSDT

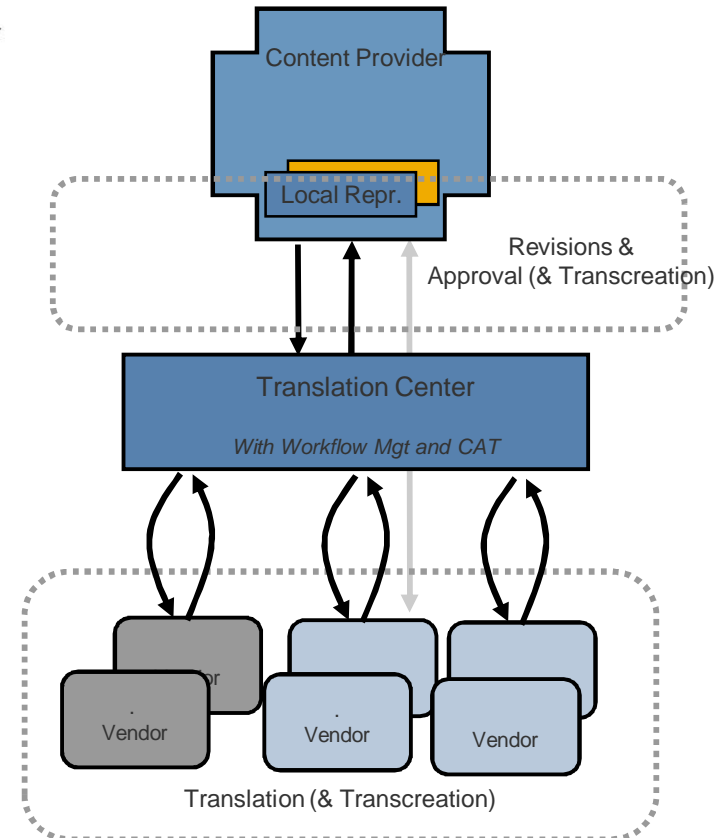
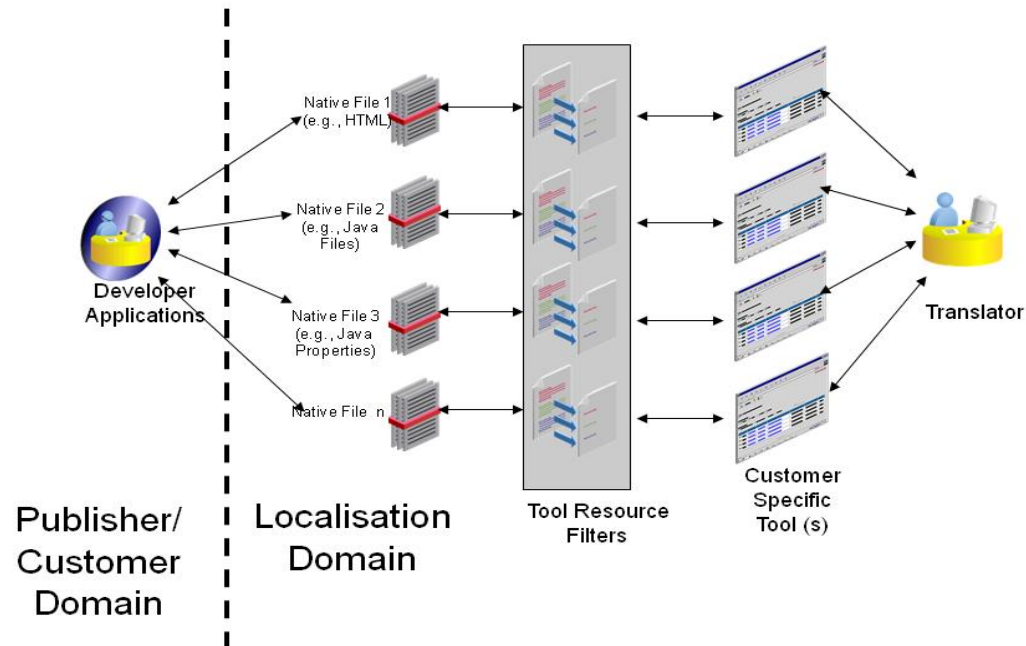
## Technology

- Editing applications
- Coupling technologies (eg. WebDAV)
- Language-related capabilities of all involved technologies (eg. related to search)
- Infrastructure for professional service delivery translation

## Resources

- Roles and responsibilities during project
- Roles and responsibilities for productive solution
- Roles and responsibilities wrt. SLS core activities
- Resources for volume business (eg. additional headcount)

# The Real World III



## Formats



## Processes



## Costs

\$\$\$¥\$¥\$¥\$¥\$¥\$¥\$  
\$\$\$¥\$¥\$¥\$¥\$¥\$¥\$

# Who is Needed for a Better World?

Developers

- Create applications for worldwide use

Process Engineers

- Follow best practices for localization

Content Producers  
and Architects

- Mark up content for worldwide use

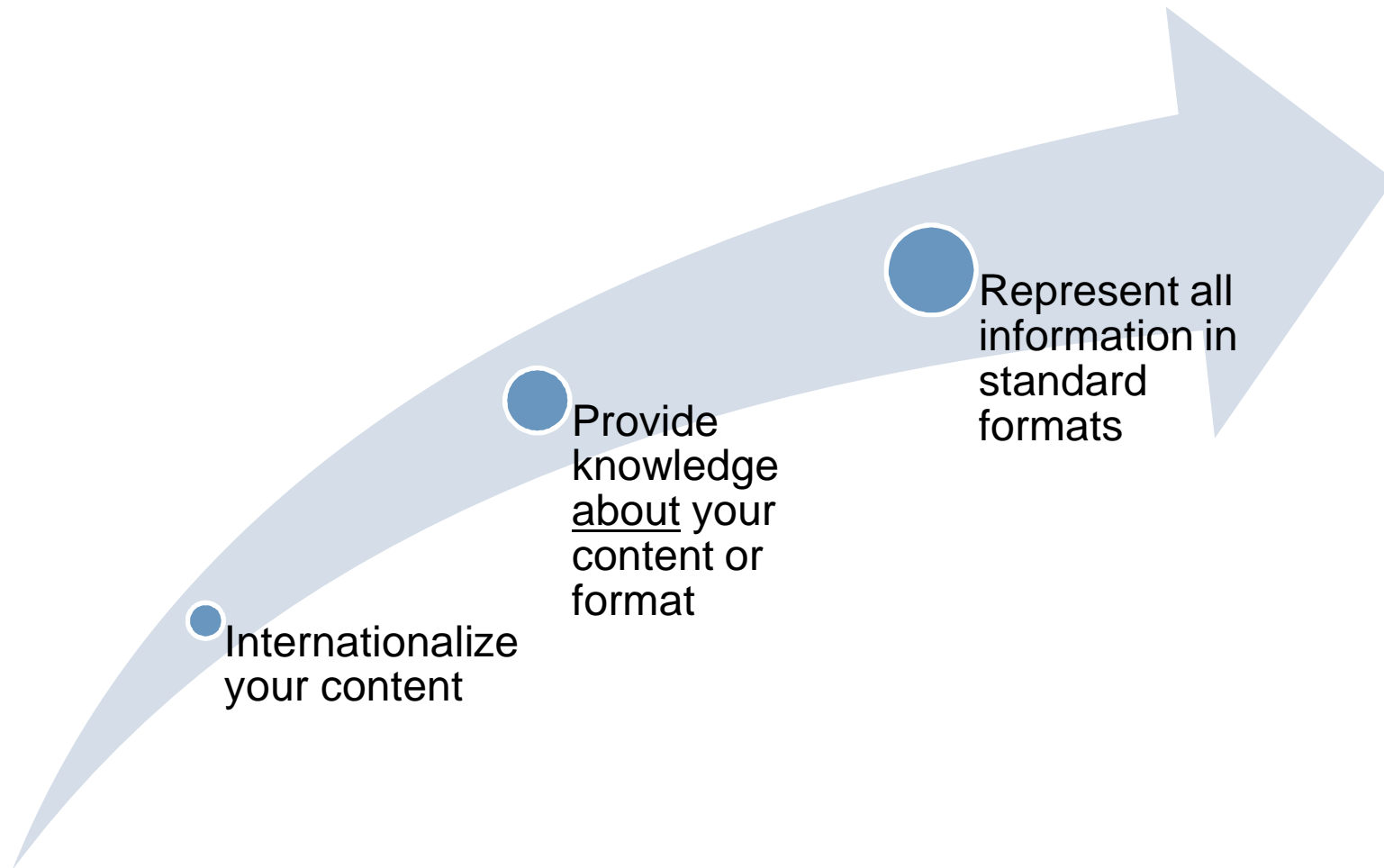
Vendors of Content-  
management tools

- Support internationalized formats and content marked up for worldwide use

Standards make their tasks easier

# What is Needed for a Better World?

A Localization Service provider may be able to help you with efficient translation of your proprietary format if you do the following



1. Challenges with Proprietary Globalization
2. **Format 1: W3C Internationalization Tag Set (ITS)**
3. Format 2: OASIS XML Localization Interchange File Format (XLIFF)
4. The Relationship between ITS and XLIFF
5. Tool: XLIFF-related converters based on ITS
6. Q&A

# FORMAT 1: W3C INTERNATIONALIZATION TAG SET (ITS)

1 Support international  
use

2 Support localization  
needs

3 Protect from translatability  
problems

4 Make meaning of tags  
easy to recognize

5 Don't disturb

# The Basic Idea by Abstraction

Say important things

- *Do not translate*

About specific content

- All *uitext* elements

In a standard way

- `its:translate="no"`
- `its:translateRule...`

# Say Important Things – ITS Data Categories

Data Category	Translate	Whether the content of an element or attribute should be translated or not
	Localization Note	Communicate notes to localizers about a particular item of content
	Terminology	Mark terms and optionally associate them with information, such as definitions
	Directionality	Specify the base writing direction of blocks, embeddings and overrides for the Unicode bidirectional algorithm
	Ruby	Provide a short annotation of an associated base text, particularly useful for East Asian languages
	Language Information	Express the language of a given piece of content
	Elements Within Text	Identify how an element behaves relative to its surrounding text, eg. for text segmentation purposes

# The Basic Idea by Example

## Local Approach

<para>

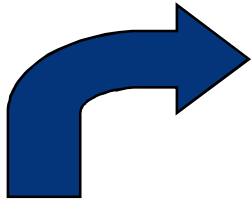
Press the

<uitext its:translate="no">START</uitext>

button to sound the horn. The

<uitext its:translate="no">MAKE-READY/ RUN</uitext> indicator flashes.

</para>



<para>

Press the

<uitext>START</uitext>

button to sound the horn. The

<uitext>MAKE-READY/ RUN</uitext>

indicator flashes.

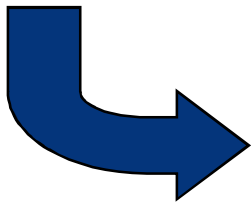
</para>

## Global Approach

<its:rules ... its:version="1.0">

<its:translateRule selector="//uitext" translate="no"/>

</its:rules>



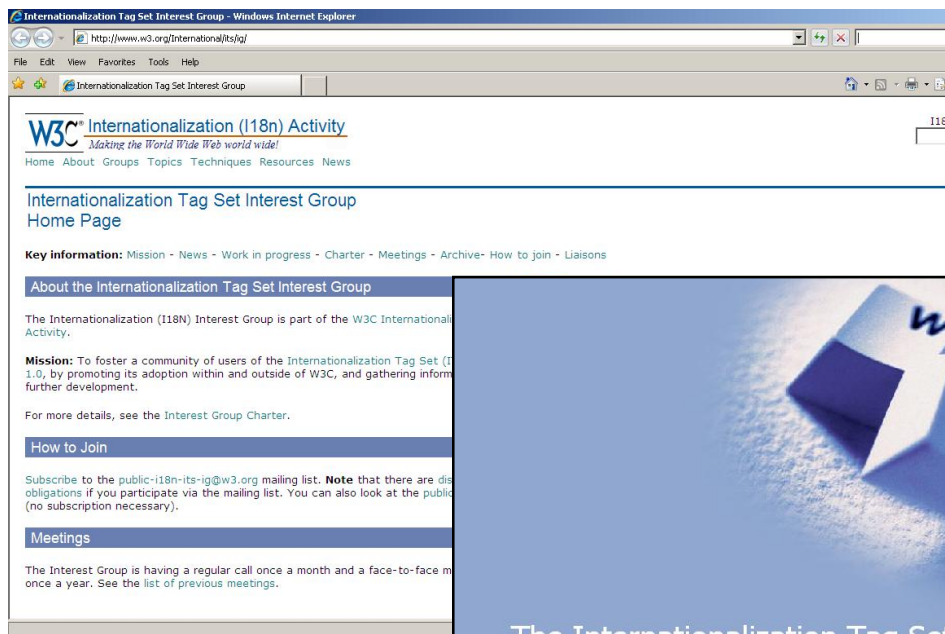
# W3C ITS – Further Information

## Specification

<http://www.w3.org/TR/its/>

W3C ITS Interest Group

<http://www.w3.org/International/its/ig/>



## Internationalization and Localization of XML: Introducing "ITS"



Christian Lieske

Sebastian Rhatz

Felix Sasaki

Slides:

<http://www.w3.org/2006/Talks/0518-xtech-its/>

## The Internationalization Tag Set

Richard Ishida  
W3C Internationalization Activity Lead

<http://www.w3.org/2006/Talks/10-Irc-its/slides/Slide0010.html>



1. Challenges with Proprietary Globalization
2. Format 1: W3C Internationalization Tag Set (ITS)
3. **Format 2: OASIS XML Localization Interchange File Format (XLIFF)**
4. The Relationship between ITS and XLIFF
5. Tool: XLIFF-related converters based on ITS
6. Q&A

# **FORMAT 2: OASIS XML LOCALIZATION INTERCHANGE FILE FORMAT (XLIFF)**

1 Unify the world

2 Recognize leveraging possibilities

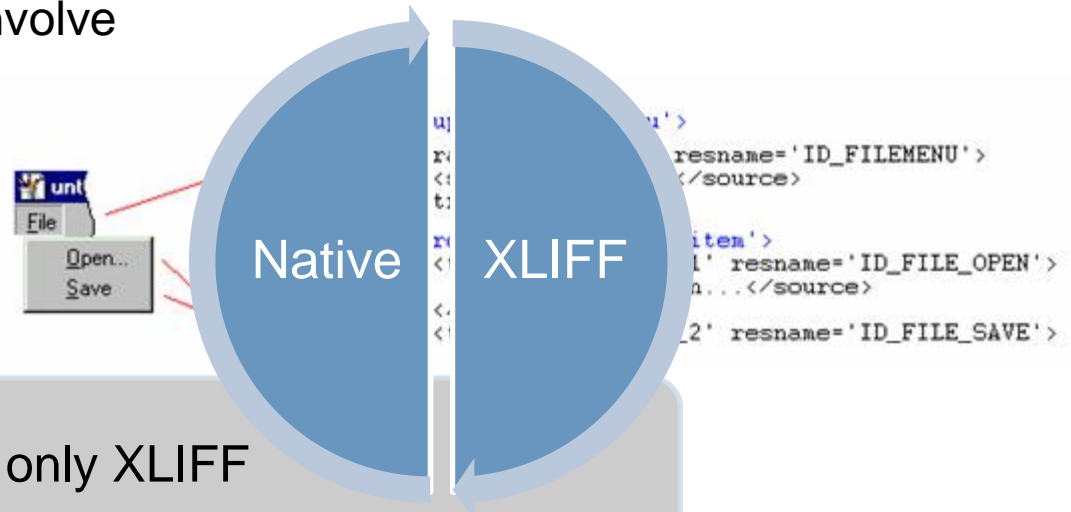
3 Have room for annotations and processing artefacts

4 Remember the way back

# The Basic Idea

XLIFF processes most of the time involve two formats ...

- Native: You don't want this.
- XLIFF: This is what you want



Identical

- There is only XLIFF

Minimalistic

- The XLIFF is mainly translatable text

Maximalistic

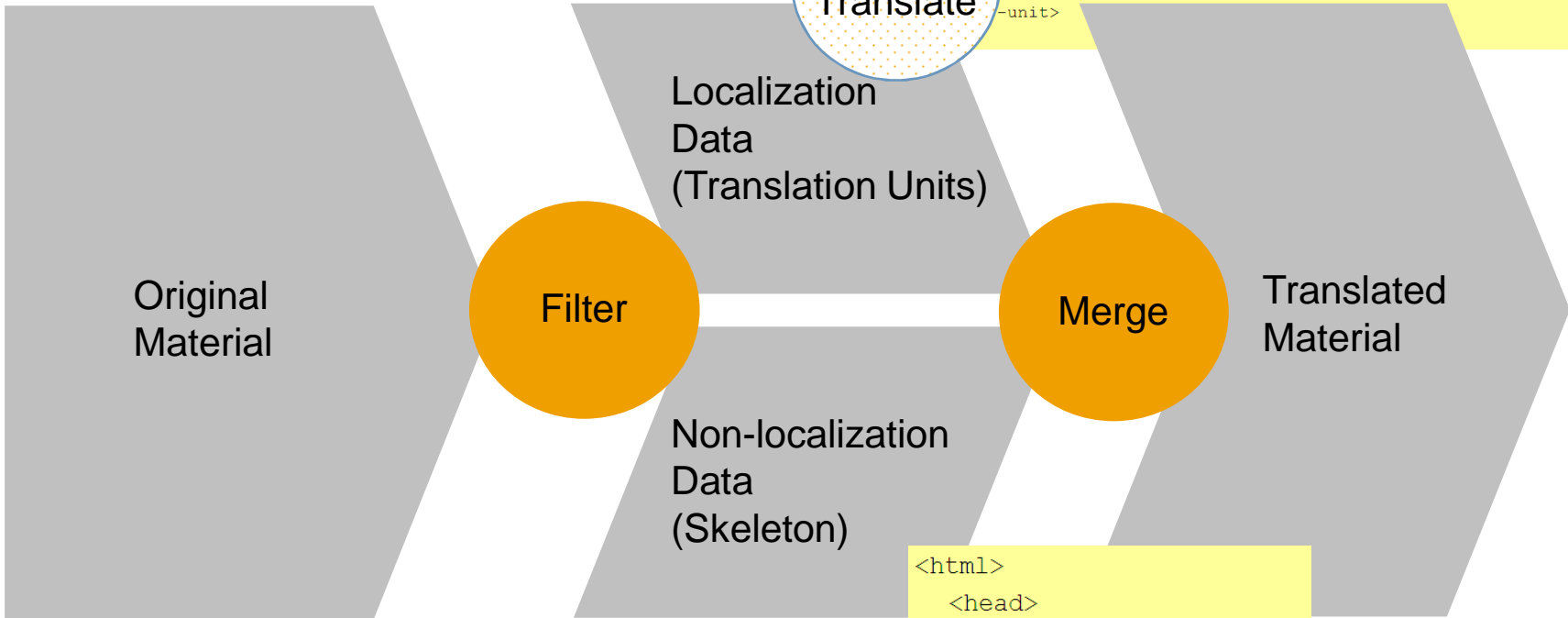
- The XLIFF is much like your original (not just the text is included)

Several so-called profiles/representation guides suggest how to map native formats to XLIFF.

# The Minimalistic Choice

```
<header>
  <skl>
    <external-file href='sample.skl' />
  </skl>
</header>
<body>
  <trans-unit id='###1###'>
    <source xml:lang='en'>Almost the Smallest HTML File</source>
  </trans-unit>
  <trans-unit id='###2###' restype='x-html-p'>
    <source xml:lang='en'>Just some stuff here to fill up space</source>
  </trans-unit>
</body>
```

Localize/  
Translate



```
<html>
  <head>
    <title>###1###</title>
  </head>
  <body>
    <p>###2###</p>
  </body>
</html>
```

# The Maximalistic Choice

Original  
Material

Filter

Localize/  
Translate

Merge

Translated  
Material

```
<body>
  <group restype='x-html-html'>
    <group restype='x-html-head'>
      <trans-unit id='1' restype='x-html-p-title' html:class='title'>
        <source xml:lang='en'>Almost the Smallest HTML File</source>
      </trans-unit>
    </group>
    <group restype='x-html-body'>
      <trans-unit id='2' restype='x-html-p'>
        <source xml:lang='en'>Just some stuff here to fill up space</source>
      </trans-unit>
    </group>
  </group>
</body>
```

# The Basic Idea by Example

An XLIFF document can capture anything needed for a localization project:

- **Localizable objects** (e.g. text strings) in source and target languages
- **Supplementary information** (e.g. glossaries, or material to recreate the original format)
- **Administrative information** (e.g. workflow data)
- **Custom data** (e.g. initialization information for tools)

```
<?xml version="1.0"?>
<xliff version="1.0">
  <file original="myFile.ext" source-language="la" tool="myTool" datatype="text">
    <header/>
    <body>
      <trans-unit id="1">
        <source xml:lang="la">Verba.</source>
      </trans-unit>
    </body>
  </file>
</xliff>
```

## 1 Simple meta data (resource type, content format) for individual texts

```
<body>
<trans-unit datatype="plaintext" restype="x-SAP-callToAction" id="idcallToAction73651136">
  <source>See Yelena's Impossible Story</source>
  <target>See Yelena's Impossible Story</target>
</trans-unit>

<trans-unit datatype="plaintext" restype="x-SAP-subtitle" id="idsubtitle73891017">
  <source>CELEBRATE<g id="1" ctype="x-lb"/>ORIGINALITY</source>
  <target>CELEBRATE<g id="1" ctype="x-lb"/>ORIGINALITY</target>
</trans-unit>

<trans-unit datatype="plaintext" restype="x-SAP-label" id="idlabel81951424">
  <source>REPLACE THIS</source>
  <target>REPLACE THIS</target>
</trans-unit>

<trans-unit datatype="plaintext" restype="x-SAP-label" id="idlabel73242496">
  <source>PORSCHE</source>
  <target>PORSCHE</target>
</trans-unit>

<trans-unit datatype="plaintext" restype="x-SAP-label" id="idlabel73241304">
  <source>Y-3</source>
  <target>Y-3</target>
</trans-unit>

<trans-unit datatype="x-xml-mapped" restype="x-SAP-subtitle" id="idsubtitle73310744">
  <source>FUSION OF<g id="1" ctype="x-lb"/>STYLE AND<g id="2" ctype="x-lb"/>PERFORMANCE</source>
  <target>FUSION OF<g id="1" ctype="x-lb"/>STYLE AND<g id="2" ctype="x-lb"/>PERFORMANCE</target>
</trans-unit>

<trans-unit datatype="plaintext" restype="x-SAP-label" id="idlabel73427784">
  <source>SHOP PERFORMANCE</source>
  <target>SHOP PERFORMANCE</target>
</trans-unit>
```

## 2 Pointers to placeholders in skeleton

## 3 Mapping of inline markup

## 4 Multiple files (XLIFF as packaging format)

## 4 Phase information

```
<?xml version="1.0" encoding="UTF-8" ?>  
- <xliff version="1.0">  
- <file source-language="en" target-language="fr" datatype="database"  
  original="SO 2003 database">  
- <header>  
- <phase-group>  
  <phase phase-name="Post-to-BerlitzIT" process-  
    name="translation" Company-name="special-olympics"  
    tool="berlitzit" date="2001-10-31" job-id="btz0002" Contact-  
    name="peter" contact-email="peter.reynolds@berlitz.ie"  
    contact-phone="353-1-2021280" />  
- </phase-group>
```

## 5 Multi-source leveraging

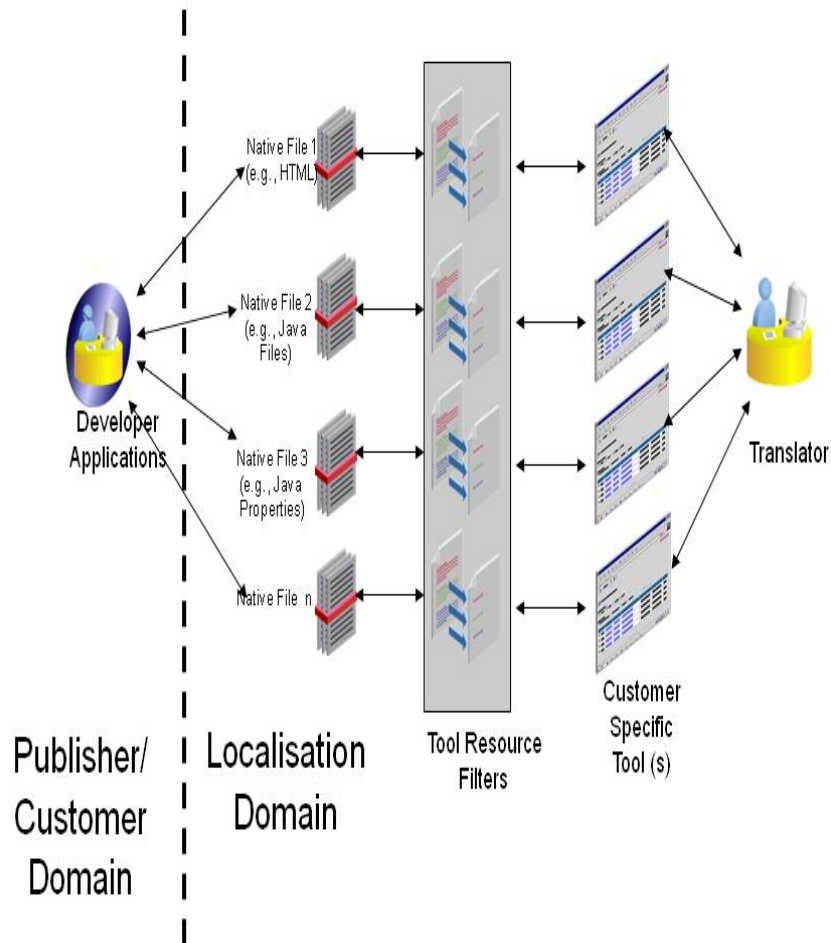
```
- <trans-unit id='1'>  
  <source xml:lang='en'>The text</source>  
  <alt-trans quality-match='high'  
    origin='MTsystem'>  
    <target xml:lang='fr'>Le texte</target>  
  </alt-trans>  
  <trans-unit id="1">  
    <source>First sentence. Second sentence.</source>  
  </trans-unit>  
  <seg-source>  
    <mrk mtype="seg" mid="1">First sentence.</mrk>  
    <mrk mtype="seg" mid="2">Second sentence.</mrk>  
  </seg-source>  
  <target>  
    <mrk mtype="seg" mid="1">Translated first sentence.</mrk>  
    <mrk mtype="seg" mid="2">Translated second sentence.</mrk>  
  </target>  
</trans-unit>
```

## 6 Pre-Segmentation

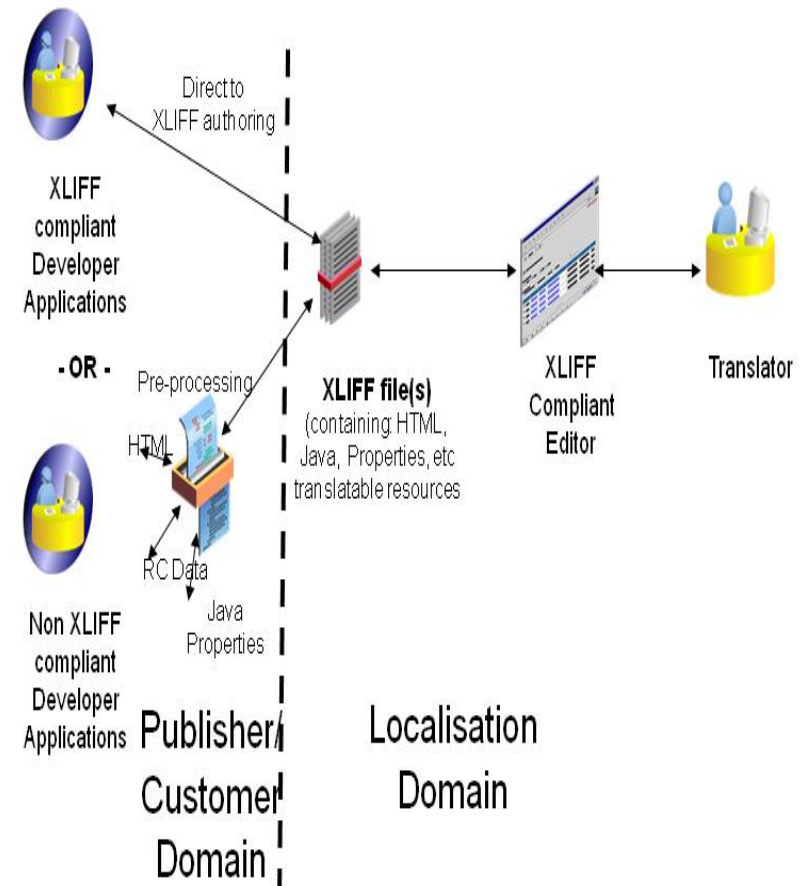
## 7 Extension mechanisms

# XLIFF, a Format to the Rescue for All (1/2)!

## Without XLIFF



## With XLIFF



# XLIFF, a Format to the Rescue for All (2/2)!

XLIFF is for anyone in localisation. It is not (just) a Localisation Industry Standard.

## Localization Customer

- Single format for adjunct processing (e.g. quality control in terms of spell checking).
- Less dependency on vendors which are able to work with special formats.
- Tighter control on what goes to localization (pre-filtering of what to translate or not).
- Must develop own tools, use one customised for them or use standard formats
- ...

## Localization Service Provider

- Single format for adjunct processing (e.g. quality control in terms of spell checking).
- Less dependency on specific localization tools (reduced training need).
- Complexity of many different formats for different customers
- Expertise may become superficial
- ...

## Tool Vendor

- Focus on development of core functionality rather treatment of source format.
- All advantages of XML-based processing
- Allows use of existing tools in new contexts.
- ...



# OASIS XLIFF – Further Information

Specification (note: in addition representation guides exist)

<http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html>

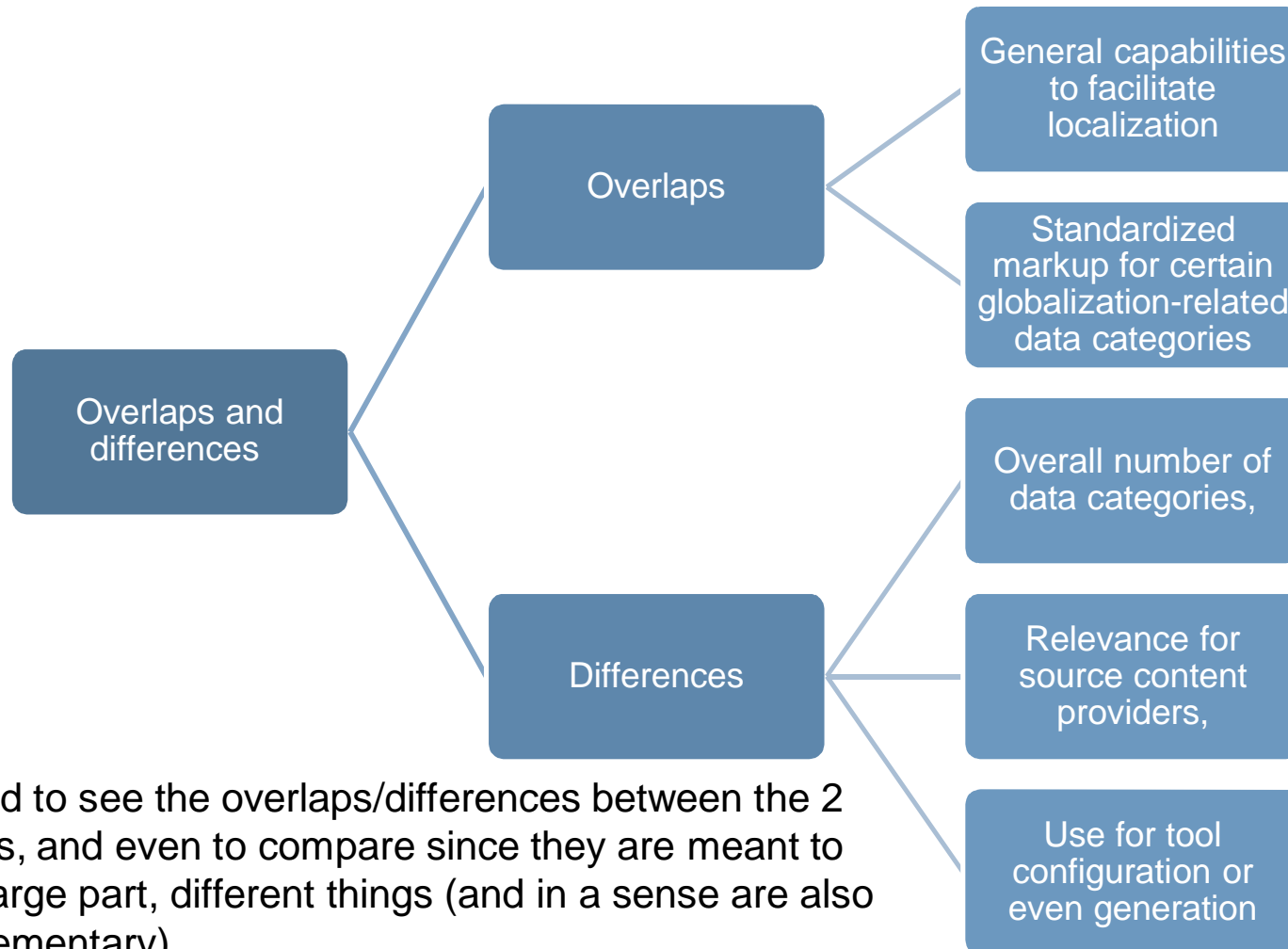
Technical Committee

[http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=xliff](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff)

1. Challenges with Proprietary Globalization
2. Format 1: W3C Internationalization Tag Set (ITS)
3. Format 2: OASIS XML Localization Interchange File Format (XLIFF)
4. **The Relationship between ITS and XLIFF**
5. Tool: XLIFF-related converters based on ITS
6. Q&A

# THE RELATIONSHIP BETWEEN ITS AND XLIFF

# Relationships – Overlaps and Differences



## Remarks:

1. It's hard to see the overlaps/differences between the 2 formats, and even to compare since they are meant to do in large part, different things (and in a sense are also complementary).
2. The list of overlaps and differences is incomplete
3. Sometimes, it is hard to decide whether one is looking at an overlap or a difference

ITS and XLIFF overlap since they

Facilitate globalization

Provide markup for globalization

Complement existing markup

Hold promise for microformats

Path the way for the future multilingual Web

# Overlap – Facilitating Globalization

## Internationalization

- ITS (via the XML Internationalization Best Practices: <http://www.w3.org/TR/xml-i18n-bp/>)

## Localization/Translation

- XLIFF: through standard container element for translatable content, (leveraged) translations, and limited set of inline markup, ...
- ITS: through formal representation of information relevant for processes

Ideal formal representation of your content related to standard formats includes

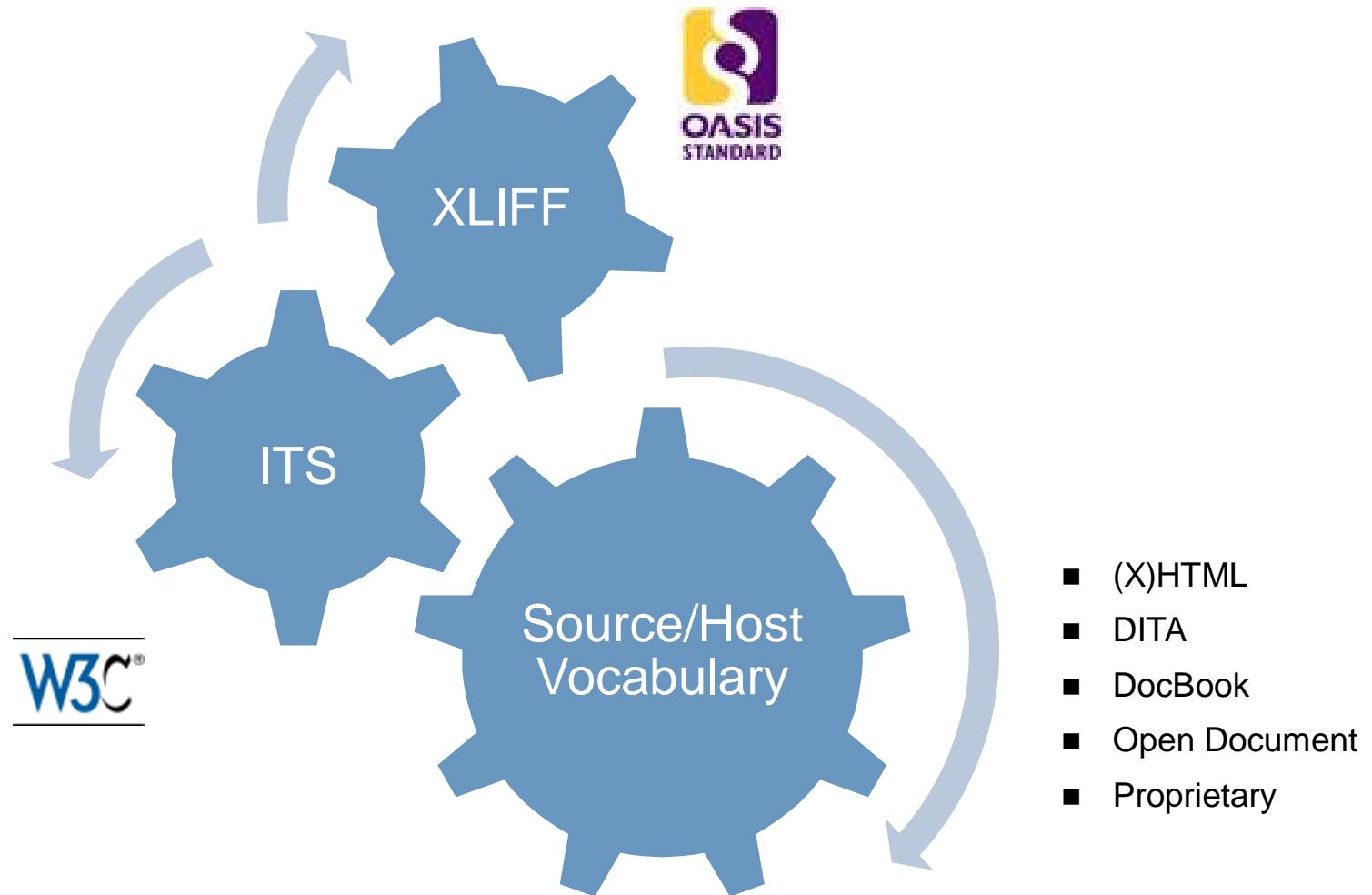
1. DTD or XSD
2. ITS rules file
3. XLIFF

# Overlap – Providing Markup for Globalization

	ITS	XLIFF
<b>Translate</b>	translate, translateRule	source
<b>Localization Note</b>	locNote, locNoteRule	note (with annotates)
<b>Terminology</b>	term, termRule	mrk (with mtype)
<b>Directionality</b>	dir, dirRule	
<b>Ruby</b>	ruby, rubyRule	
<b>Language Information</b>	xml:lang, langRule	xml:lang
<b>Elements Within Text</b>	withinText, withinTextRule	ept, bpt, g

Caveat: This side-by-side view needs further work/verification.

# Overlap – Complement Formats (1/2)



Complementation requires attention to maximize benefits, clarify “division of labor” and reuse of markup.

Defaults, and inheritance may be challenging.

# Overlap – Complement Formats (2/2)

## W3C ITS

```
<its:rules
xmlns:its="http://www.w3.org/2005/11/its"
its:version="1.0">

<its:translateRule selector="//uitext"
translate="no" />

</its:rules>

<para>Press the

  <uitext its:translate="no">START</uitext>

  button to sound the horn. The

  <uitext its:translate="no">MAKE-READY/
RUN</uitext>

  indicator flashes. </para>
```

## OASIS XLIFF

```
<table
xmlns:xlf=""urn:oasis:names:tc:xliff:document:
1.2"">

  <title>Android Memory </title>

  <tbody><row><entry>

    <xlf:trans-unit id="Z123">

      <xlf:source xml:lang="en">one two
three</xlf:source>

      <xlf:target xml:lang="de">ein zwei
drei</xlf:target>

    </xlf:trans-unit>

  </entry></row></tbody>

</table>
```

[http://docs.oasis-open.org/xliff/v1.2/os/xliff-core.html#Struct\\_Embedding](http://docs.oasis-open.org/xliff/v1.2/os/xliff-core.html#Struct_Embedding)  
Adapted from XLIFF Whitepaper [http://www.oasis-open.org/committees/download.php/3110/XLIFF-core-whitepaper\\_1.1-cs.pdf](http://www.oasis-open.org/committees/download.php/3110/XLIFF-core-whitepaper_1.1-cs.pdf)

# Overlap – Promising Microformats (1/2)

Microformats are based on the fact that XHTML and HTML allow for semantics via the CLASS and REL attributes

Existing microformats cover domains such as contact information, geographic coordinates, calendar events, and social relationships

Since semantics is explicitly encoded, and represented in a standardized way, automated processing becomes easier

Several browsers realize support for microformats (possibly via extension/plugin)

Example (geographic coordinates):

```
The birds roosted at  
<span class="geo">  
  <span class="latitude">52.48</span>,  
  <span class="longitude">-1.89</span>  
</span>
```

Test your imagination (ITS) ...

```
<span class="tel">  
  <span class="type">Home</span>  
  <span class="tel">  
    <span class="its-translate-no">+1.415.555.1212</span>  
  </span>  
</span>
```

## Overlap – Promising Microformats (2/2)

Discussion related to microformats for ITS has already been started

<http://lists.w3.org/Archives/Public/public-i18n-its-ig/2009Jul/0000.html> and  
[http://docs.google.com/View?id=dch8cn8g\\_20hrxhkmd8](http://docs.google.com/View?id=dch8cn8g_20hrxhkmd8)

Implementation for the "Translate" data category

1. Class attributes (example: global rules/the *its:translateRule* element)

```
<span class="translateRule">  
  <span class="selector"> //code</span>  
  <span class="translate">no</span>  
</span>
```

2. EBMF notation

ITSMF = itsprefix [translate] [terminology] [localizationNote]

[directionality]

itsprefix = "its,,

translate = "-translate-" ("yes" | "no")

terminology = "-term" ["-termInfoRef:" IRI] ; IRI production from RFC 3987

localizationNote = ...

# Overlap – Pathing the Way (1/4)

The World Wide Web amongst others is a collection of interrelated resources

One mechanism for expressing a relationship between a resource like a static HTML page is the LINK element

The LINK element's REL attribute describes the relationship between the current document and the anchor specified by the HREF attribute

Thus, the LINK element conveys relationship information that may be rendered or interpreted by user agents in a variety of ways

Example: CSS Stylesheet

```
<LINK href="mystyle.css" rel="stylesheet" type="text/css" />
```

Test your imagination ...

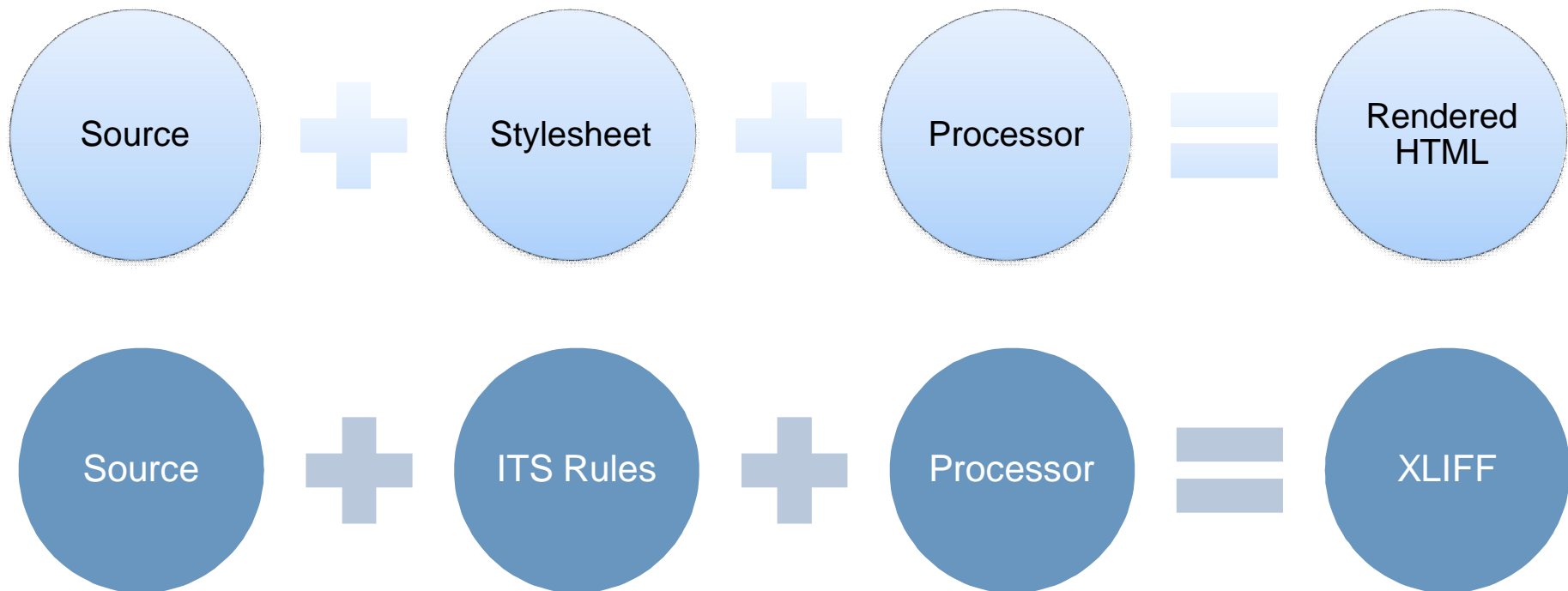
```
<LINK href="myi18nrules.its" rel="i18nrules" type="application/xml+its" />
```

# Overlap – Pathing the Way (2/4)

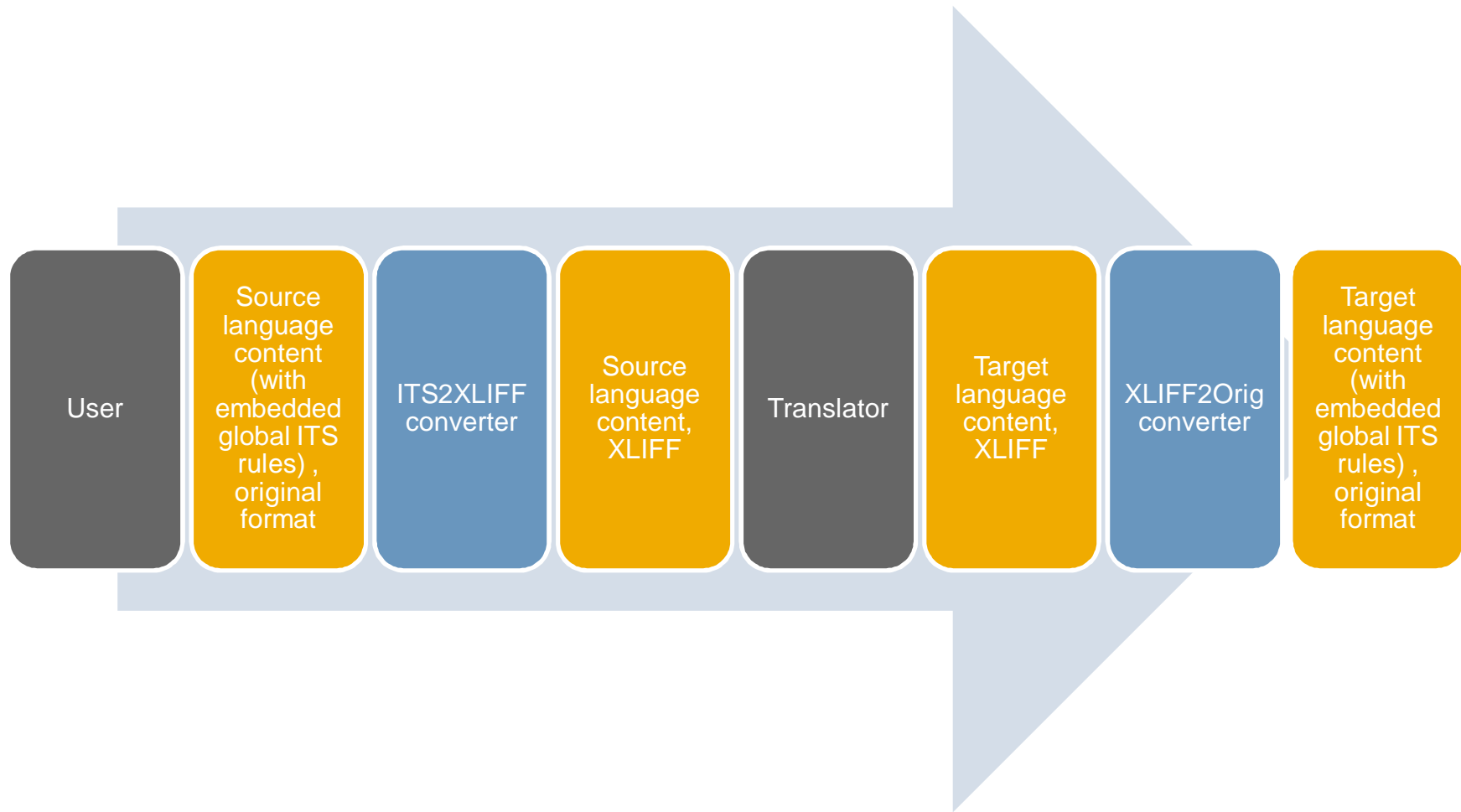
A user agent could use ITS rules for converting content into XLIFF.

Discussion related to a MIME-type for ITS has already been started

<http://lists.w3.org/Archives/Public/public-i18n-its-ig/2009Jul/0011.html>

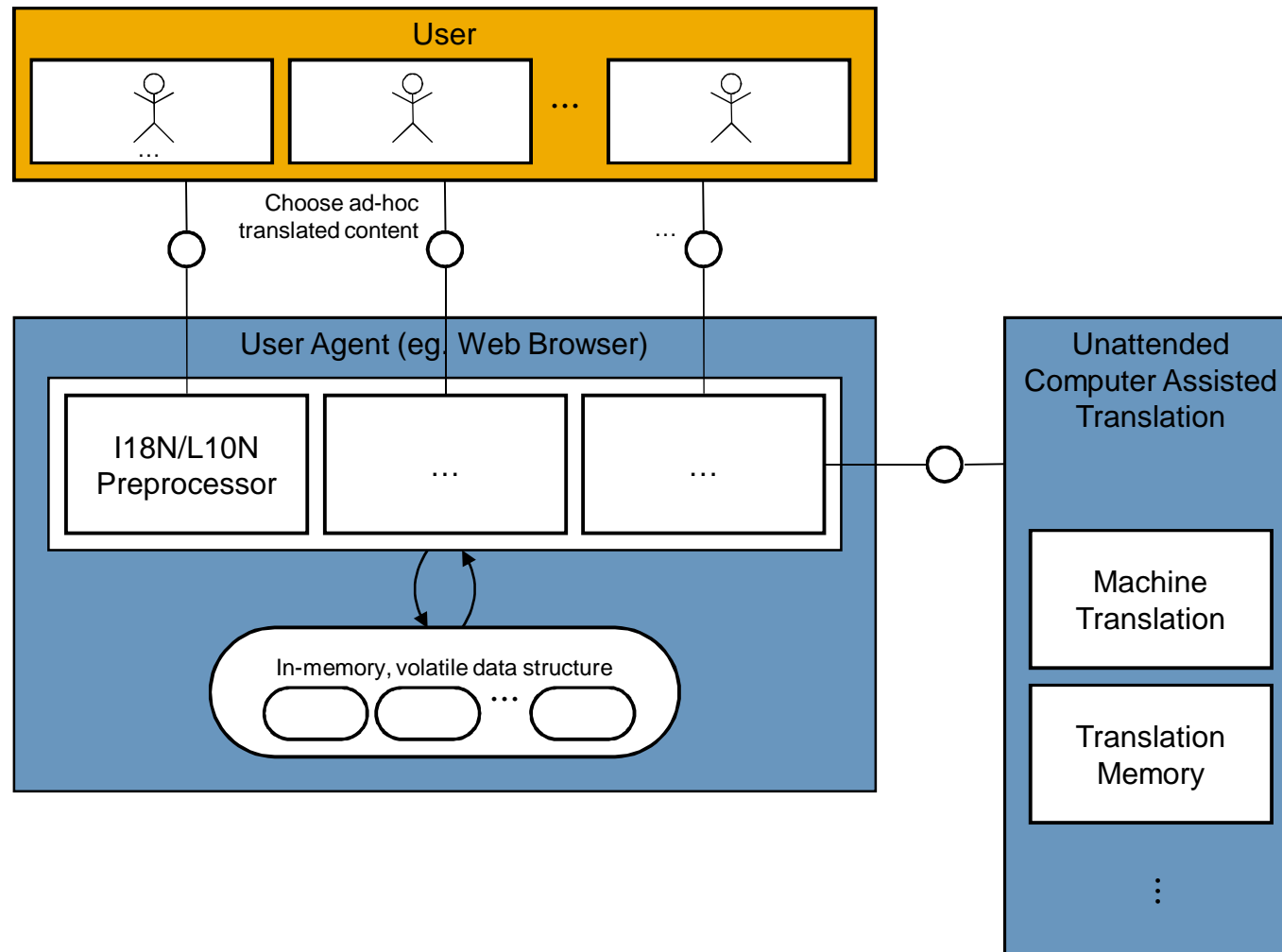


# Overlap – Pathing the Way (3/4)



# Overlap – Pathing the Way (4/4)

Internationalization and Localization for distributed resources based on user clients interpreting ITS and XLIFF!



ITS and XLIFF differ with regard to

Relationship to content payload

Relevance for source content providers

Use for tool configuration

Use for tool generation

# Difference – Relationship to Content Payload

## XLIFF File

Source  
content

100%  
matches  
from  
Translation  
Memory

Other  
matches  
from  
Translation  
Memory

Reference  
(eg. extract  
from  
Terminology  
Database)

Results from  
Machine  
Translation

....

## Example: SDL Trados Studio 2009

SDL Trados Studio 2009 is the latest incarnation of the well known SDL Trados Translation Memory and Project Management technology.

In the context of SDL Trados 2009, there is no longer a need to create proprietary configurations when working with proprietary or rarely used XML file formats. Rather – thanks to support of ITS – configuration happens automatically and on the fly.

If an XML file has ITS markup embedded, or references ITS-rules, identification and processing of translatable content happens automatically. To be specific so-called „parser rules“ are generated on-the-fly. This differs from XML files without ITS markup, where parser rules have to be created in advance.

Aside: Initial ITS support has focussed on the the ITS data categories „translate“ and „withinText“. Additional support is under discussion.

# Difference – Tool Configuration (2/5)

**Project Settings - Project 1**

**File type information**

File type name: XML: W3C ITS Compliant

File type icon: assembly://Sdl.FileTypeSupport.Native.Xml\_1\_1/Sdl.FileTypeSupport.Native.Xml.FilterDefinition.ico

Settings file: C:\Documents and Settings\d025418\My Documents\SDL Trados Studio\Projects\Project 1\File Types\XML\_ITS\_1\_0.sdl

File type identifier: XML: ITS 1.0 v 1.1.0.0

**File names**

Filename must match the following regular expression:

Name of individual document: ITS XML Document

Name of document category: ITS XML Documents

File dialog wildcard expression: \*.xml

**Description:**

XML document utilizing local rules, embedded rules and/or explicit rule files following the W3C Internationalization Tag Set (ITS) standard (version identification and processing of localizable content).

**SDL Trados Studio - Project 1**

File View Project Tools Help

Open Document... New Project... Open Package... Batch Tasks... File details layout

Name	Words	Status	Progress	Size	Usage	File Type Identifier	Path
BM_2_00_en_US2.xml	n/a			274 KB	Translatable	XML: ITS 1.0 v 1.1.0.0	\

Files

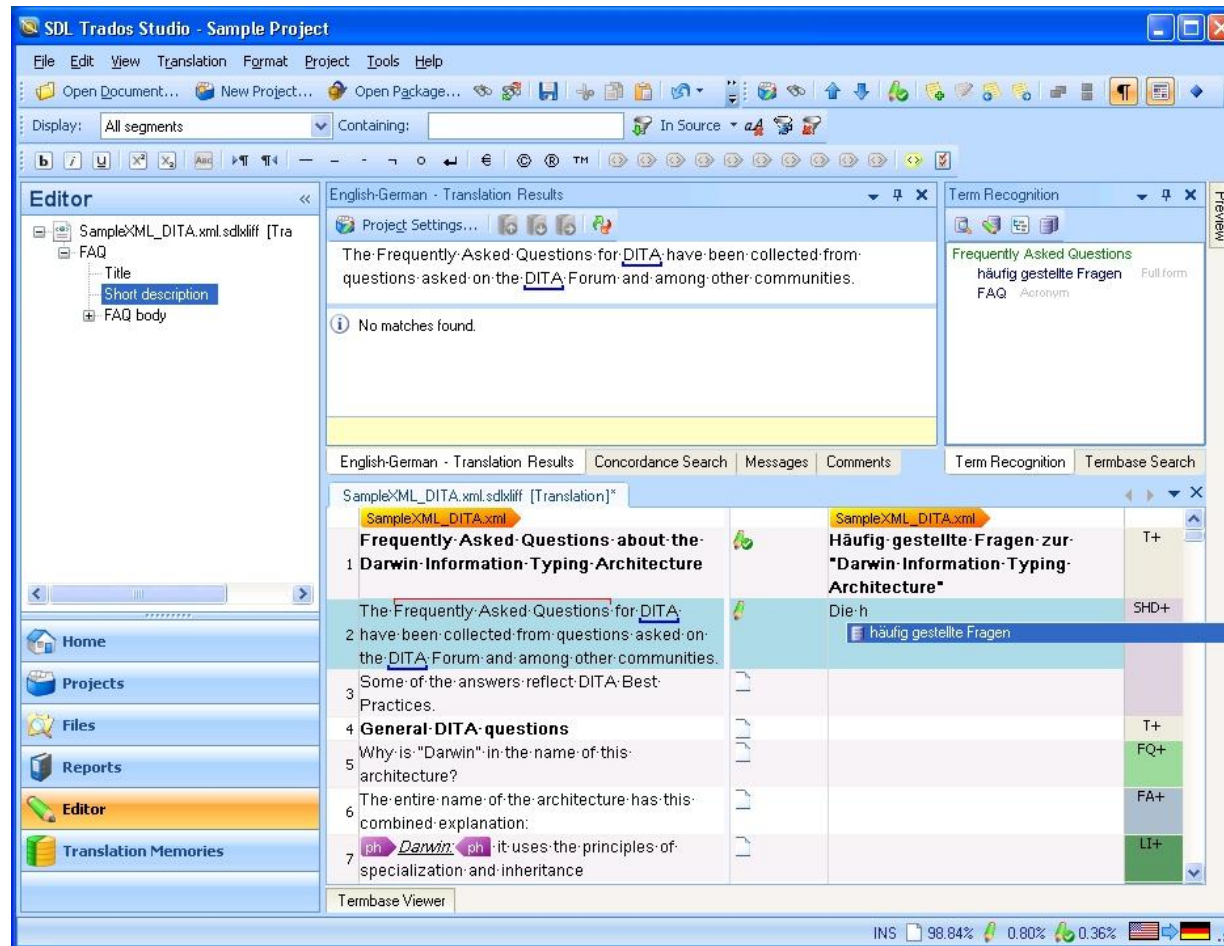
English (United States)

Project folders

- Project 1
- My Tasks
- Sent Tasks

# Difference – Tool Configuration (3/5)

XML file using ITS in the Editor. (DITA XML file shipping with Studio in the sample project)



# Difference – Tool Configuration (4/5)

## Example: Yves Savourel, ENLASO Corporation

<http://okapi.translate.com/Utilities/ITSTest.aspx>

**ITS Test Implementation**

This page allows you to test some of the [XML Internationalization Tag Set \(ITS\)](#) rules. See the [How to Use This?](#) section for information.  
Last update: Jun-21-2006, 9:02am MDT.

**IMPORTANT: This is an early BETA implementation of the Okapi XML Filter and its support for ITS. You may (will!) experience problems with some files.**

Enter the XML source (with ITS markup if needed):

```
<!-- Example: DITA -->
<concept id="myConcept" xml:lang="en-us">
  <prolog>
    <its:rules xmlns:its="http://www.w3.org/2005/11/its" its:version="1.0">
      <its:translateRule selector="//@alt" translate="yes"/>
      <its:translateRule selector="//*[@translate='no']" translate="no"/>
      <its:translateRule selector="//*[@translate='no']/descendant-or-self::*/*" translate="no"/>
      <its:translateRule selector="//*[@translate='yes']" translate="yes"/>
      <its:translateRule selector="//*[@translate='yes']/descendant-or-self::*/*" translate="yes"/>
    </its:rules>
  </prolog>
  <title>Concept Title</title>
  <conbody>
    <p>Palouse horses<fn>\ palouse horse is the same as an Appalosa.<image href="appalosa.png" alt="Appalosa horse"/></fn> have spotted coats.</p>
    <p translate="no">Palouse horses<fn>\ palouse horse is the same as an Appalosa.<image href="appalosa.png" alt="Appalosa horse"/></fn> have spotted coats.</p>
    <p translate="no">Palouse horses<fn translate="yes">\ palouse horse is the same as an Appalosa.<image href="appalosa.png" alt="Appalosa horse"/></fn> have spotted coats.</p>
  </conbody>
</concept>
```

Extract to XLIFF   Pseudo-Translate   Test Results   Select an example...

# Difference – Tool Configuration (5/5)

## Example: acrolinx IQ

acrolinx IQ™ amongst other enables a variety of configurable linguistic checks

Configuration related to identification of content to be checked is done by means of Context Segmentation Definition (CSD) files

The following ITS rules can be easily used to create a corresponding CSD.

```
<its:rules version="1.0"
xmlns:its="http://www.w3.org/TR/2007/REC-its-20070403">
  <its:translateRule selector="//*" translate="no" />
  <its:translateRule selector="//source
translate="yes" />
  <its:withinTextRule withinText="yes" selector="g |
x | bx | ex | ph | bpt | ept | it" />
</its:rules>
```

```
# Sample CSD for XLIFF files which hold content
to be checked in "source" elements
# Inline elements have been defined according
to http://docs.oasis-open.org/xliff/v1.2/os/xliff-
core.html#Specs_Elem_Inline
# Other dimensions of the CSD (e.g.
"inline_headline_elements_used") have not be
addressed yet

headline_elements=source
inline_elements=g, x, bx, ex, ph, bpt, ept, it, sub
inline_headline_elements_used=
skipped_elements=header
filetype=xml
```

# Difference – Tool Configuration/Generation

Example: Felix Sasaki, Christian Lieske (details and demo in next section)

<http://fabday.fh-potsdam.de/~sasaki/its/>

ITS2XLIFF generation version 0.6

The W3C Internationalization Tag Set (ITS) defines data categories and their implementation as a set of elements and attributes called the Internationalization Tag Set (ITS). ITS is designed to be used with schemas to support the internationalization and localization of schemas and documents. Please take a look at [link collection](#) for more ITS-related information (amongst others presentations, and references to implementations).

Aside: The W3C ITS Interest Group gives ample opportunities to get involved.

The XML Localization Interchange File Format (XLIFF) gives any documentation or software provider a single interchange file format that can be understood by any localization provider. This page allows you to generate XLIFF 1.2 from XML files for which W3C ITS rules is available (local markup or global rules). Currently, only the ITS ITS 1.0 Transtate is supported.

The servicefunctionality behind the page is the ITS General Decorator, an XSLT-based ITS processor which decorates an input document with information for ITS data categories. To be specific, the service works along the lines of the XLIFF Extraction and Merging.

If you want to test the service, you can for example use the samples in the ITS Test Suite.

Feedback is very welcome - please contact Felix Sasaki.

Warning: This is experimental, a stable version is yet to come.

Last Modified: 09/24/2009 11:10:34

For the generation of an XLIFF file, use this form:

XML File

For the integration of translated content into the original XML file, use this form:

XLIFF File for conversion back to XML

Layout based on YAML

1. Challenges with Proprietary Globalization
2. Format 1: W3C Internationalization Tag Set (ITS)
3. Format 2: OASIS XML Localization Interchange File Format (XLIFF)
4. The Relationship between ITS and XLIFF
5. **Tool: XLIFF-related converters based on ITS**
6. Q&A

# **TOOL: XLIFF-RELATED CONVERTERS BASED ON ITS**

XLIFF Roundtrip (non-ITS) – Bryan Schnabel

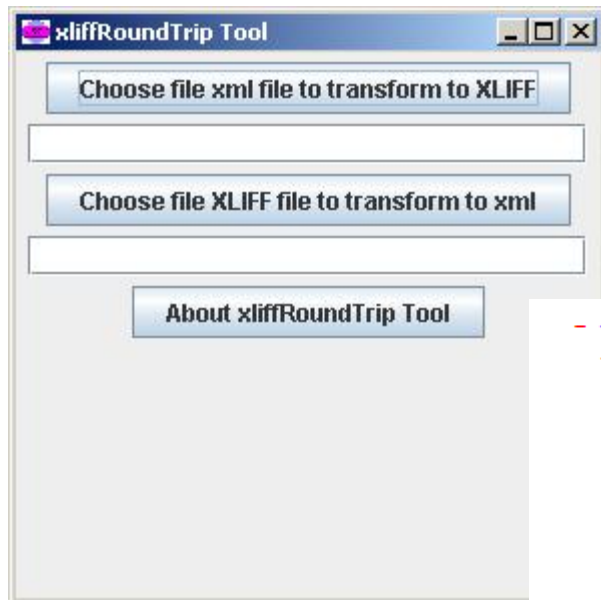
Okapi –The Okapi Framework team” (7 developers, on 3 continents and 4 time zones)

General Decorator – Felix Sasaki, Christian Lieske

1. Disclaimer: The authors and programmers or their employers shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of the programs.
2. Due to XLIFF's flexibility (for example wrt. skeleton files)
  - The XLIFF created by the converters differs
  - The XLIFF created by the individual converters implements only one possible representation (cf. the distinction between minimalistic and maximalistic XLIFF)

# XLIFF Roundtrip Tool

## Graphical User Interface



## Web Resources

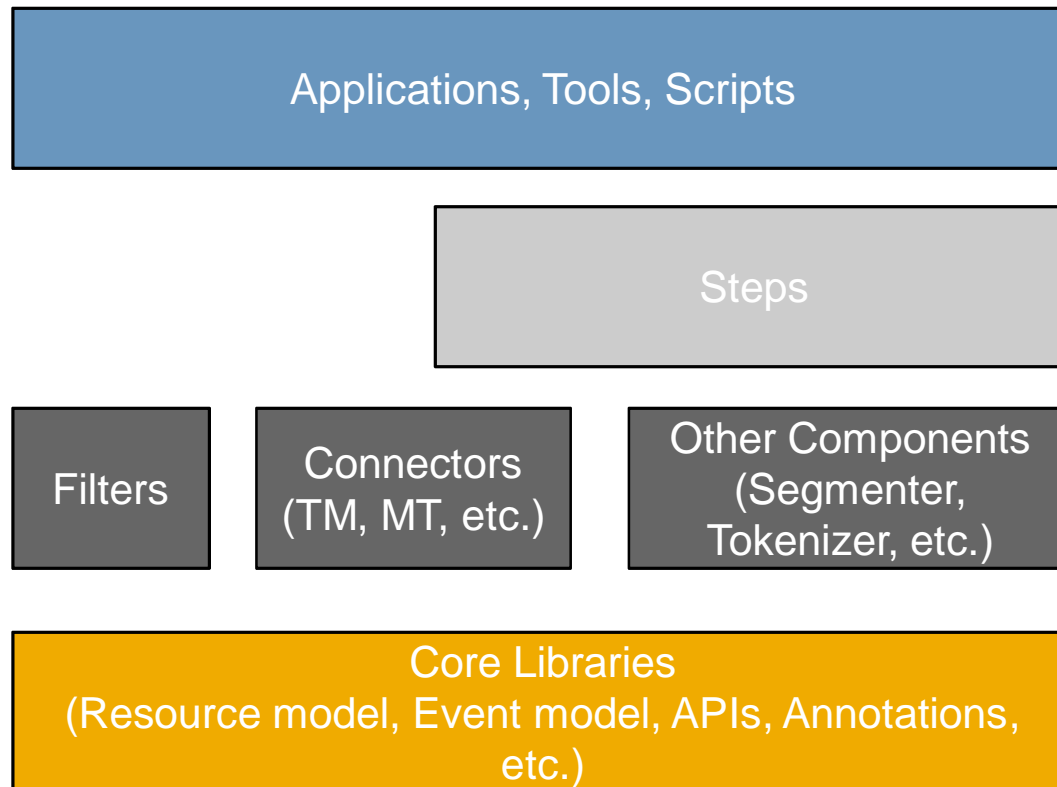
<http://sourceforge.net/projects/xliffroundtrip/>

```
- <body>
- <group id="N65537axmarktext-0">
- <trans-unit id="para-1">
- <source>
  Press the
  <g id="uitext-2">START</g>
  button to sound the horn. The
  <g id="uitext-3">MAKE-READY/ RUN</g>
  indicator flashes.
</source>
- <target>
  Press the
  <g id="uitext-2">START</g>
  button to sound the horn. The
  <g id="uitext-3">MAKE-READY/ RUN</g>
  indicator flashes.
</target>
</trans-unit>
</group>
</body>
</file>
</xliff>
```

Resulting XLIFF  
(no markup for  
elements which are not  
translation-relevant)

Set of libraries and components for writing translation- and localization-related tools

## Web Resources



Downloads, source, bugs, etc:  
<http://code.google.com/p/okapi>

Help and documentation:  
<http://okapi.opentag.com>

Users Group:  
<http://code.google.com/p/okapi>

## Application: Rainbow (source XML with its associated ITS parameter file)

The screenshot displays a Windows Internet Explorer browser window showing the source XML of a document. The XML structure is as follows:

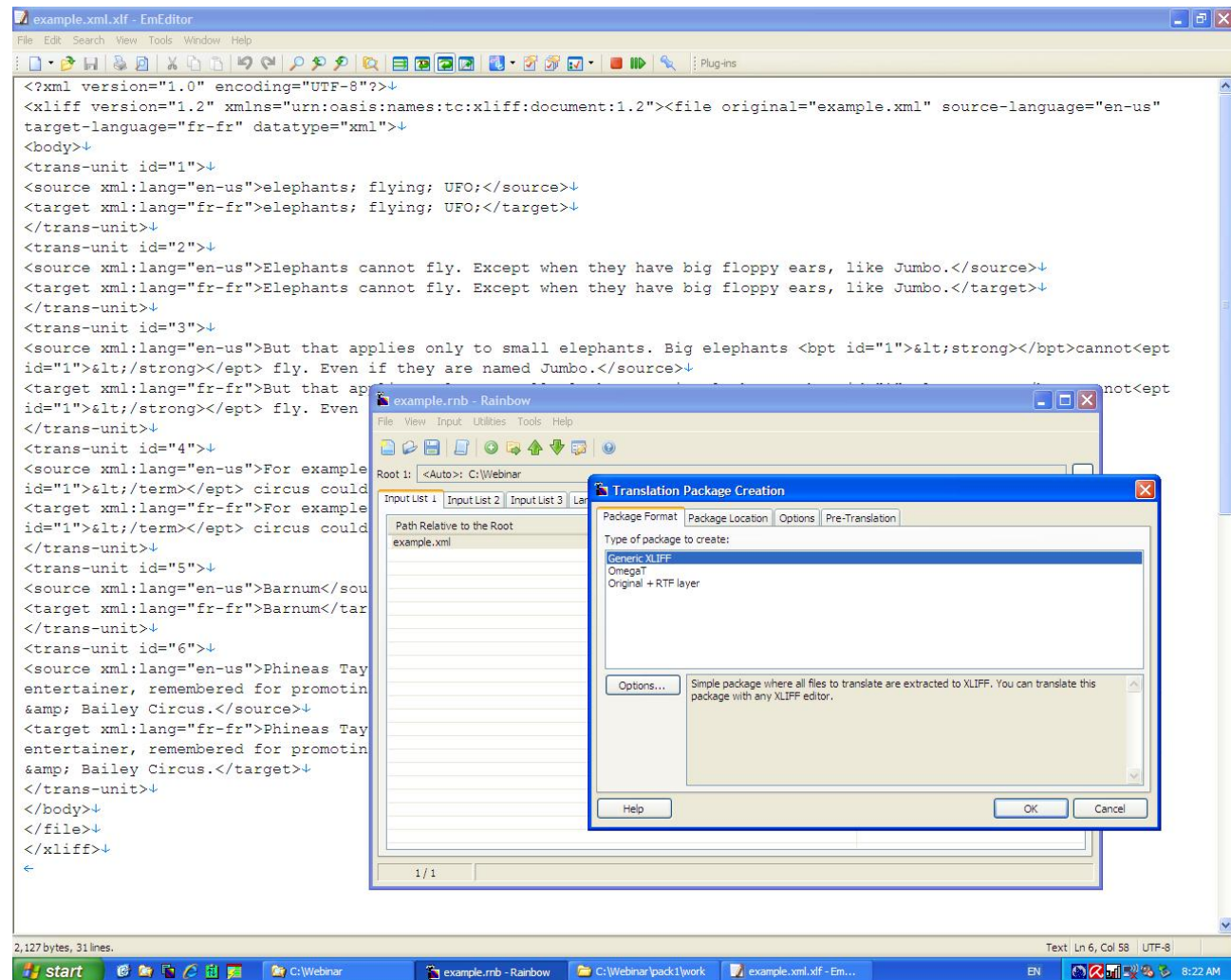
```
<?xml version="1.0" ?>
<doc>
  <header>
    <author>Andrezj Artimus Prakash</author>
    <lastUpdate>20091019T13:23:04Z</lastUpdate>
    <keywords>elephants; flying; UFO;</keywords>
  </header>
  <comment>Some non translatable comment</comment>
  <body>
    <p>Elephants cannot fly. Except when they have big floppy ears, like Jumbo</p>
    <p>But that applies only to small elephants. Big elephants cannot fly. Even if they are named Jumbo.</p>
    <p>For example the Jumbo from the Barnum</p>
    <comment>Some internal comments that should not be translated</comment>
    <termDef id="barnum">
      <term>Barnum</term>
      <def>Phineas Taylor Barnum (July 5, 1810 - July 6, 1899) was an American naturalist, inventor, and showman who organized the first American circus and for founding the American Museum of Natural History.</def>
    </termDef>
  </body>
</doc>
```

Overlaid on the browser window is a dialog box titled "Filters Parameters (okf\_xml@myDoc)". The dialog contains the following ITS rules:

```
<?xml version="1.0" encoding="UTF-8"?><its:rules xmlns:its="http://www.w3.org/2005/11/its" xmlns:itsx="http://www.w3.org/2008/12/its-extensions" xmlns:xlink="http://www.w3.org/1999/xlink" version="1.0">
<!-- See ITS specification at: http://www.w3.org/TR/its -->
<its:translateRule selector="//doc/header" translate="no"/>
<its:translateRule selector="//doc/header/keywords" translate="yes"/>
<its:translateRule selector="//comment" translate="no"/>
<its:withinTextRule selector="//strong//term" withinText="yes"/>
<its:withinTextRule selector="//termDef/term" withinText="no"/>
<its:termRule selector="//*[name()='termDef']/term" term="yes" termInfoPointer="id(@ref)"/>
</its:rules>
```

The dialog also shows a list of input files, including "okf\_xml@myDoc", and options for source and target encodings.

## Rainbow with resulting XLIFF



## Design decisions:

1. Zero installation
2. For XLIFF: internal skeleton file, derivation of *restypes* from elements names, first sprint for *translation* data category
3. For ITS: initially, only support for „translate“ data category

## Web Resources

<http://fabday.fh-potsdam.de/~sasaki/its/>

ITS2XLIFF generation version 0.6

The W3C Internationalization Tag Set (ITS) defines data categories and their implementation as a set of elements and attributes called the Internationalization Tag Set (ITS). ITS is designed to be used with schemas to support the internationalization and localization of schemas and documents. Please take a look at link collection for more ITS-related information (amongst others presentations, and references to implementations).

Aside: The W3C ITS Interest Group gives ample opportunities to get involved.

The XML Localization Interchange File Format (XLIFF) gives any documentation or software provider a single interchange file format that can be understood by any localization provider. This page allows you to generate XLIFF 1.2 from XML files for which W3C ITS rules is available (local markup or global rules). Currently, only the ITS ITS 1.0 Translate is supported.

The service functionality behind the page is the ITS General Decorator, an XSLT-based ITS processor which decorates an input document with information for ITS data categories. To be specific, the service works along the lines of the XLIFF Extraction and Merging.

If you want to test the service, you can for example use the samples in the ITS Test Suite.

Feedback is very welcome - please contact F. Sasaki:

Warning: This is experimental, a stable version is yet to come.

Last Modified: 09/24/2009 11:10:34

For the generation of an XLIFF file, use this form:

XML File

For the integration of translated content into the original XML file, use this form:

XLIFF File for conversion back to XML

Layout based on YAMIL

# General Decorator (2/2)

```
<xliff version="1.2">
- <file datatype="xml" original="unknown" source-language="en" date="2001-12-17T09:30:47.0Z">
  - <header>
    - <skl>
      - <internal-file>
        - <![CDATA[
          <text>
            <para xmlns:its="http://www.w3.org/2005/11/its"
              its:dummy_for_its_1="http://www.w3.org/2005/11/its">
              <xref id="d1t6"/>
              <uitext its:dummy_for_its_1="http://www.w3.org/2005/11/its" its:translate="no">
                START
              </uitext>
              <xref id="d1t11"/>
              <uitext its:dummy_for_its_1="http://www.w3.org/2005/11/its" its:translate="no">
                MAKE-READY/ RUN
              </uitext>
              <xref id="d1t16"/>
            </para>
          </text>
        ]]>
      </internal-file>
    </skl>
    <tool tool-name="IGD-2-XLIFF" tool-id="IGD-2-XLIFF" tool-version="0.1" tool-company="World Wide Web Consortium - Internationalization Tag Set Interest Group; OASIS XLIFF TC" />
  </header>
- <body>
  - <trans-unit xmlns:its="http://www.w3.org/2005/11/its" datatype="plaintext" restype="x-para" id="d1t6">
    <source>Press the</source>
    <target>Press the</target>
  </trans-unit>
  - <trans-unit xmlns:its="http://www.w3.org/2005/11/its" datatype="plaintext" restype="x-para" id="d1t11">
    <source>button to sound the horn. The</source>
    <target>button to sound the horn. The</target>
  </trans-unit>
  - <trans-unit xmlns:its="http://www.w3.org/2005/11/its" datatype="plaintext" restype="x-para" id="d1t16">
    <source>indicator flashes.</source>
    <target>indicator flashes.</target>
  </trans-unit>
</body>
</file>
</xliff>
```

1. Challenges with Proprietary Globalization
2. Format 1: W3C Internationalization Tag Set (ITS)
3. Format 2: OASIS XML Localization Interchange File Format (XLIFF)
4. The Relationship between ITS and XLIFF
5. Tool: XLIFF-related converters based on ITS
6. Q&A

## Q&A

# Contact

Let us know if you have any questions, ideas etc.

# Thank you!

Christian Lieske [christian.lieske@sap.com](mailto:christian.lieske@sap.com)  
Felix Sasaki [felix.sasaki@fh-potsdam.de](mailto:felix.sasaki@fh-potsdam.de)  
Bryan S. Schnabel [bryan.s.schnabel@tektronix.com](mailto:bryan.s.schnabel@tektronix.com)  
Yves Savourel [ysavourel@translate.com](mailto:ysavourel@translate.com)