

Takeaways from the First International XLIFF Symposium



Christian Lieske (SAP AG)
September 27, 2011 – Warsaw

The 2nd International XLIFF Symposium



Presenter



Christian Lieske

**SAP Language Services
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, the Unicode Consortium, the European Commission and elsewhere
- Degree in Computer Science with focus on Natural Language Processing and Artificial Intelligence

The First International XLIFF Symposium 2010 (1/2)

Limerick, Ireland – Context of yearly conference run by the Localization Research Center (LRC)

More than 70 attendees from all around the globe; 18 contributions

Researchers, tool providers, language service buyers and providers, ...

1ST XLIFF INTERNATIONAL SYMPOSIUM 22 September 2010, Limerick, Ireland

The 1st International XLIFF Symposium took place in Limerick on 22nd September 2010. This event was the main pre-conference activity for the [LRC XV: 15th Annual Internationalisation and Localisation Conference](#) organised by the [LRC](#). This was the first ever solo XLIFF event and it was a complete success in terms of participants, presenters and quality of discussions.

The symposium brought together specialists, tools providers, developers, TC members and researchers to discuss and share their experience with [XLIFF](#) (XML Localisation Interchange File Format). New trends were presented, implementation cases were demonstrated, and the future of XLIFF and its new version 2.0 was discussed.

Intense, heated discussions followed each presentation and the high level of audience participation was surprisingly positive as well as enriching.

Photos of the event



Thanks to Eoin Ó Conchúir, Jasenka Wasele, Christian Loeke and Naoto Nishio for the photos. If you would prefer that we remove a photo of you, simply email luca.mora@ul.ie.

Keynote Speaker



Bryan Schmechel – Chair of the XLIFF Technical Committee, and XML Information Architect for Tektronix Inc.

"XLIFF's Place in the Community"

A community is a group sharing common characteristics and interests, perceiving itself as distinct from the larger society within which it exists. Ours is the Localisation Community. It consists of many disciplines. In it we find Localisation Service

<http://www.localisation.ie/xliff/>

The First International XLIFF Symposium 2010 (2/2)

Type	Presenter(s)	Title
Plenary Sessions	Reinhard Schäler (Welcome)	
	Bryan Schnabel (Keynote speaker)	XLIFF's Place in the Community
	Asanka Wasala, Dag Schmidtke, Reinhard Schäler	XLIFF and LCX Format: A Comparison
	Asgeir Frimannsson, Christian Lieske	Next Generation XLIFF: Simplify - Clarify - and Extend
	JoAnn Hackos, Bryan Schnabel, Rodolfo Raya	DITA and XLIFF: A Great Marriage
Panels	Moderator: Christian Lieske; Panellists: Asgeir Frimannsson, Christian Lieske, Stefan Pries, Friedel Wolff	Minimal and modular XLIFF
	Moderator: David Filip; Panellists: Gábor Ugray, Lorcan Ryan, Heiko Rölke, Dag Schmidtke, David Filip	XLIFF metadata throughout workflows
Parallel Sessions	Niall Murphy	The route to XLIFF adoption in Oracle
	Felix Sasaki, Christian Lieske	XLIFF and ITS: A secret marriage
	Martin Beuster, Gábor Ugray	XLIFF as a potential TM exchange format?
	Steve Dept, Andrea Ferrari, Britta Upsing, Heiko Rölke	Case study: XLIFF in a large-scale international OECD-study
	Micah Bly	XLIFFs in Theory and in Reality
Short Presentations	Shirley Coady	Increasing Quality in the Translation Process through the Integration of XLIFF with Advanced Leveraging Translation Memory and Terminology
	Friedel Wolff	XLIFF from a volunteer's point of view
	Derek Coffey	GlobalSight as a case study in the deployment and use of XLIFF as an interchange format
	Thomas Vackier	The problem of extensibility in XLIFF 1.2
	Jörg Schütz, Alexandra Weissgerber	Localising Business Process Repository Content
Posters	Naoto Nishio, Ian O'Keeffe, J.J. Collins	Taxonomy of localisation services with XLIFF

Takeaways – General Input to the Work of the XLIFF TC

Acceptance and a desire for conformance/compliance guidelines

Vast support for modular and minimal architecture

Wide range, differences of opinions on what exactly should be in the *core module(s)*

Pleas for backward compatibility

Extensibility needs to be examined and perhaps diminished

Better mechanism to include binary information should be explored

Wide range of usage scenarios

Takeaways – Critical Voices

No tool supports all of what's in XLIFF 1.2 (implied that XLIFF 1.2 is too feature-rich, and unwieldy)

Get XLIFF 2.0 to the community rather sooner than later

Genesis of the Detailed Takeaways (1/2)

OASIS XLIFF Technical Committee (TC) has gone through all of the contributions


Distilled input for three different questions related to XLIFF

What is the status quo related to implementation, and adoption?

Which ideas or recommendations exist for enhancements, or future versions?

Which general observations can be made?

Genesis of the Detailed Takeaways (2/2)

OASIS  **XLIFF Wiki** [Login](#)

Consolidated Takesaways from First XLIFF Symposium/

[FrontPage](#) [RecentChanges](#) [FindPage](#) [HelpContents](#) [Consolidate...F Symposium](#)

[Immutable Page](#) | [Info](#) [More Actions:](#)

This page is used by the XLIFF Technical Committee to gather takeaways from the [First XLIFF Symposium](#). Deadline for submitting takeaways is 15 October 2010.

Type	Presenter(s)	Title	Takeaway (Please Prefix by your Name)
Plenary Sessions	Reinhard Schäler (Welcome)		
	Bryan Schnabel (Keynote speaker)	XLIFF's Place in the Community	
	Asanka Wasala, Dag Schmidtke, Reinhard Schäler	XLIFF and LCX Format: A Comparison	<p><i>Observations related to XLIFF Implementation Status Quo</i></p> <p>Christian:</p> <ol style="list-style-type: none">1. tool vendors prefer proprietary formats2. implementation by tool vendors could be improved: more features, more recent version (namely 1.2), more transparency wrt. supported features, higher accuracy wrt. supported features3. open source community prefers PO format <p><i>Ideas/Recommendations related to XLIFF</i></p> <p>Christian</p> <ol style="list-style-type: none">1. create object model (similar to for example XML DOM; to allow programmatic access that is transparent to the underlying data format)2. include mechanism to represent or refer to original data (e.g. binary data for accurate, secure representation of User Interfaces)3. include mechanism to represent or refer to validation rules4. include mechanism for flexible annotation for human consumption (might be an enhanced "note" element) of any element of attribute of the representation5. include mechanism for flexible (typed) annotation for machine processing of any element of attribute of the representation6. create container-based representation format (similar to for example the ODF or OOXML zip formats)7. simplify and clarify (so that for example there is a canonical representation, not a choice between different representations)8. include mechanism for flexible annotation of referenced resources

<http://wiki.oasis-open.org/xliff/Consolidated%20Takesaways%20from%20First%20XLIFF%20Symposium>

Categories of the Contributions

Status Quo

Implementation
and
Adoption/Usage

- Survey – Where are we?
- Analysis – Why did we get there?

Future

Ideas and
Recommendations

- General
- Simplification
- Clarification
- Extension

Credits: Asgeir Frimannsson

Status Quo: Where are we?

Proprietary extensions

Intransparent feature sets

Buggy implementations

Lack of interoperability and flavouritis

Low number of implementations

Few open source offerings (Virtual notable example)

Support in commercial tools

Use in large scale projects

Staggered deployment

Status Quo: Why did we get there?

Representations are not orthogonal

- The same thing can be represented in different ways

Processing requirements are not clearly defined

- Tools don't always know how to modify XLIFF

Extensibility features have been abused

- A proprietary approach rather than a standard mechanism was used

Future Versions: General

Do not reinvent the wheel

- If for example a suitable approach for general annotations does already exist – for example as a trade industry standard – than strongly consider using it

Acknowledge that new approaches are available for describing resources/providing meta data

- Example: the Resource Description Format from the W3C

Try to stay backwards compatible

Establish stricter rules for the use of extensions

Take a look at all of the extensions that the tool makers have enacted, and at the really existing XLIFF files

- Use this as a clue for what to put into the core for a future XLIFF version

Define clear conformance rules

Future Versions: Simplification

Modularize the XLIFF specification and schema in such a way that implementers/users can pick and choose what you really need to implement an XLIFF profile

- Example: Annotations for human consumption aren't needed everywhere. Accordingly, some implementers may choose to exclude the "human annotation" module from the profile they implement

Define a schema that defines a minimalistic XLIFF

Provide a modular specification, and customized, role-specific user guides

Explain mechanisms to minimize markup/tags

Future Versions: Clarification (1/2)

Processing requirements currently are not always obvious

- Attribute values that tools have to change

Provenance of XLIFF's data categories and values is not traceable

- *restype* attribute

Different representations are possible

- inline markup

Not always indicated, what's deprecated

Future Versions: Clarification (2/2)

Possible phases

- Extraction/Filtering, Constraint Setting (e.g. for related to maximal length), Internationalization, Automated Linguistic Processing, Human Translation, Localization, Reviewing, Inclusion of reviewing results, Workflow Events, Tool-specific Events, Technical Quality Assurance checks (e.g. XML validation), Packaging

Possible data categories and data category clusters

- Payload (unsegmented or segmented content, translations), String length constraints (min. or max. length), Resource type (e.g. different User Interface controls – label, button, ...), Inlines („ph“, „x“, ...), Identifiers (for processing), Names (ie. the identifier used in the native format – key in a Java property file ...), Notes (e.g. explanations and other admonitions), Internationalization (e.g. „translate“), Domain/subject area, Relationships (e.g. between strings belonging to the same User Interface menu), Creation (e.g. generator, and creation date)

Future Versions: Extension (1/2)

Supporting technology

- Object model (similar to for example the XML DOM) to allow programmatic access that is transparent to the underlying data format
- Container-based representation format (similar to for example the ODF or OOXML zip formats)
- Open source library that allows creation and modification of XLIFF

Mechanisms to represent, or refer to original data (e.g. binary data for accurate, secure representation of User Interfaces)

Representation of/reference to validation rules (e.g. in Schematron)

Flexible annotations

- For any elements or attribute
- For human consumption or machine processing
- Allow a version history for the annotations
- For both core XLIFF payload (unsegmented content, segments, sub-segments, source and target) and referenced resources (like linked glossaries)

Mechanism for meta data after acceptance of "alt-trans"

Future Versions: Extension (2/2)

Mechanisms for changes in content (e.g. successive source versions), and “track changes”

Handling of linguistic variants due to plurals, gender, and the like

Project information and quality management information (id, instructions etc.)

Capability for declarations (e.g. required in the target files) related to character encoding

Mention best practices/conventions for including or referencing terminology data (might be the W3C Internationalization Tag Set mechanism to tag terms or to link to)

"concept-based" approach to translation (cf. concept-based terminology; ie. allow "source" to contain a concept identifier; related to xml:tm approach)

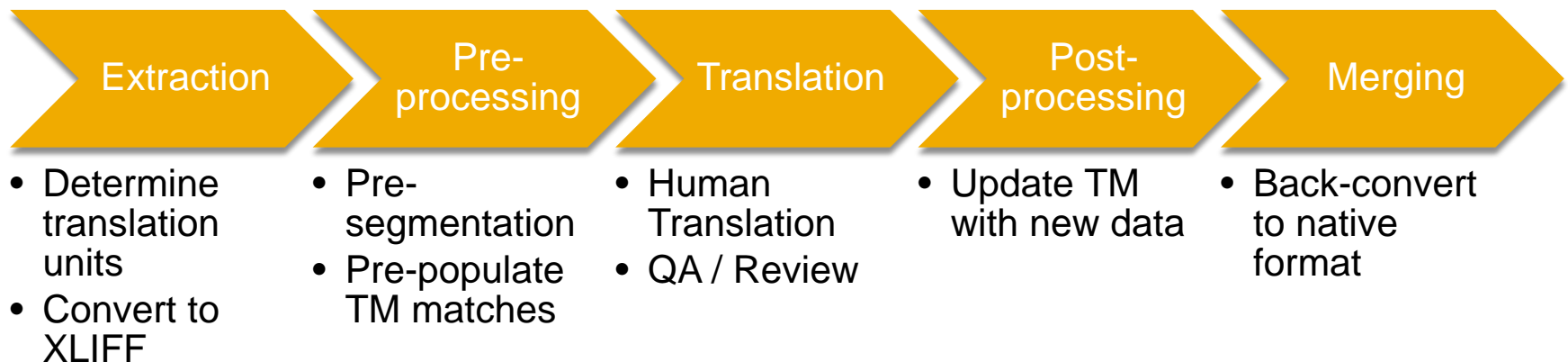
Approach authoring from a semantic modelling, not from a text writing point-of-view

Consider XLIFF as format for translation memory exchange or even general content exchange

Personal Highlights – Think Process Phases!

Do not think about XLIFF in terms of XML elements and attributes

Determine how various tools/actors manipulate XLIFF through the various stages of a localisation process



Personal Highlights – Think Process Steps!

Extraction/Filtering (including reverse ie. insertion/merging)

Constraint Setting

- Determine size requirements
- Determine character set requirements

Internationalization

- Mark "not translation relevant"
- Create note

Automated Linguistic Processing

- Segmentation
- Automatic TM Leveraging
- Machine Translation
- Terminology identification/extraction
- Automated linguistic checking (spelling, grammar, style, terminology)

Human Translation

- Creation of target text
- Display, creation or annotation of notes
- Special operations (eg. cloning)

Localization

- Adaptation of viewports
- Adaptation of URLs

Reviewing

- Insertion of annotations
- Display, creation or annotation of notes

Inclusion of reviewing results

Workflow Events

- Status updates

Tool-specific Events

- Trigger rendering

Technical QA checks

- Missing "target" after translation
- Check size requirements
- Check character set requirements

Packaging

Personal Highlights – Look at the Real World!

Payload (ie. „source“
and possibly pre-filled
„target“)

String length
constraints (min. or
max. length)

Resource type (e.g.
different User Interface
controls – label,
button, ...; overlap with
internationalization)

Inlines („ph“, „x“, ...)

Identifiers (for
processing)

Names (ie. the
identifier used in the
native format – key in
a Java property file ...)

Notes (e.g.
explanations and other
amonitions; overlap
with
internationalization)

Internationalization
(e.g. „translate“)

Domain/subject area

Relationships (e.g.
between strings
belonging to the same
User Interface menu)

Creation (e.g.
generator, and
creation date)

Personal Highlights – Look at the New World! (1/2)

Enrich “classic” with RDF

```
1 <!DOCTYPE xliif:xliif [
2   <ENTITY xsd "http://www.w3.org/2001/XMLSchema#">
3   <ENTITY vcard "http://www.w3.org/2001/vcard-rdf/3.0#">
4   <ENTITY mime "http://www.iana.org/assignments/media-types/">
5   <ENTITY domain "http://www.sap.com/mlt/domain/1.0#">
6   <ENTITY collection "http://www.sap.com/sls/collection/1.0#">
7   <ENTITY ppms "http://www.sap.com/ppms/1.0#">
8   <ENTITY lang "http://www.sap.com/mlt/lang/1.0#">
9 ]>
10 <xliif version="1.2" xmlns="urn:oasis:names:tc:xliif:document:1.2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
11   <file original="hello.txt" source-language="en" target-language="fr" datatype="plaintext">
12     <header>
13       <!-- We use RDF to describe meta data, and work with existing vocabularies/namespaces (see
14            http://dublincore.org/documents/dcmi-terms and http://www.w3.org/TR/vcard-rdf/) -->
15       <rdf:RDF xmlns:dc="http://purl.org/dc/terms#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:v="
16            http://www.w3.org/2001/vcard-rdf/3.0#">
17         <rdf:Description rdf:about="http://dtr2010/DasExecutionCompView.wdview.xlf">
18           <dc:identifier xml:lang="en-US" rdf:datatype="xsd:string">XSR Test Object 1</dc:identifier>
19           <dc:format rdf:datatype="&mime;">application/xml</dc:format>
20           <dc:language rdf:datatype="&lang;">en</dc:language>
21           <dc:collection rdf:datatype="&collection;">TEF Collection</dc:collection>
22           <dc:subject rdf:datatype="&domain;">BC</dc:subject>
23           <dc:creator rdf:nodeID="creator"/>
24           <dc:created rdf:datatype="xsd:date">2010-03-22T14:49:19+01:00</dc:created>
25           <dc:title xml:lang="en-US" rdf:datatype="xsd:string">A test file</dc:title>
26           <dc:description xml:lang="en-US" rdf:datatype="ppms;">Edition/version/release information (e.g. related to PPMS)</
27         </rdf:Description>
28         <rdf:Description rdf:nodeID="creator" rdf:type="vcard;">
29           <v:UID>d025418@sap.com</v:UID>
30         </rdf:Description>
31       </rdf:RDF>
32     </header>
33     <body>
34       <trans-unit id="hi">
35         <source>Hello world</source>
36       </trans-unit>
37     </body>
38   </file>
39 </xliif>
```

Credits: Felix Sasaki

Personal Highlights – Look at the New World! (2/2)

```
<file xmlns:dc="http://purl.org/dc/elements/1.1/">
```

```
<author property="dc:creator">Alice</author>
```

```
...
```

```
</file>
```

Credits: Felix Sasaki

Enrich similiar to RDFa

This allows you to use XLIFF pretty much as today, but you provide a bridge into other meta data worlds (for example Dublin Core). In addition, you enter the world of the Semantic Web since many tools exist for technologies such as RDF. In addition, you can already reap benefits such as improved search results: Google Rich Snippets makes use of RDFa during crawling.

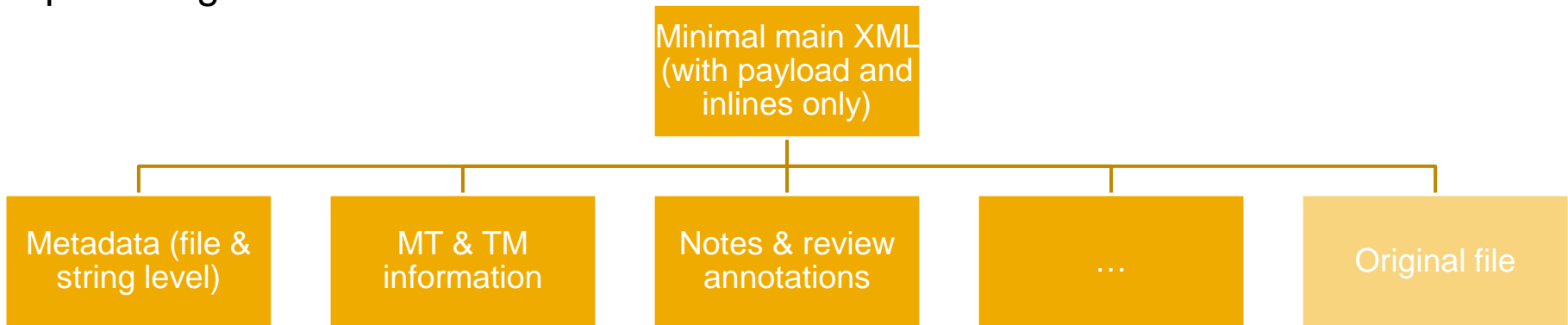
Personal Highlights – Modularize!



Personal Highlights – Package!

1. Similar to for example Open Document Format (ODF)
2. All files (except for the optional original) would be XML

Zip-Package



Credits: Stefan Pries, Asgeir Frimannsson

Conclusions

The First XLIFF Symposium provided valuable insights not just for the XLIFF community in general, but also for the XLIFF Technical Committee (TC) – the steward of the standard.

The TC intends to use the insights as guidance for ongoing work on XLIFF.

Possibly, later today, you will sense how the TC worked with the input.

Hopefully, the categories of the takeaways can be applied to today's presentations as well.

Motivated by the "Takeaways", the magazine MultiLingual (see <http://www.multilingual.com>) has made a very interesting offer for anyone interested in XLIFF: a free digital-only subscription, or a print-with-digital subscription for a reduced fee. See the details here: <http://lists.oasis-open.org/archives/xliff-comment/201107/msg00000.html>

Thank You!

Contact information:

Christian Lieske

christian.lieske@sap.com

www.sap.com

Disclaimer

All product and service names mentioned and associated logos displayed are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

This document may contain only intended strategies, developments, and is not intended to be binding upon the authors or their employers to any particular course of business, product strategy, and/or development. The authors or their employers assume no responsibility for errors or omissions in this document. The authors or their employers do not warrant the accuracy or completeness of the information, text, graphics, links, or other items contained within this material. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

The authors 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 these materials. This limitation shall not apply in cases of intent or gross negligence.

The authors have no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third-party Web pages nor provide any warranty whatsoever relating to third-party Web pages.