

W3C Internationalization Tag Set



Christian Lieske (SAP AG)
Felix Sasaki (German Center for Artificial Intelligence)

Public

W3C Workshop:
The Multilingual Web - Where Are We?
26-27 October 2010, Madrid



POLITÉCNICA
"Ingeniamos el futuro"



THE BEST-RUN BUSINESSES RUN SAP™



Agenda

1. Introduction
2. Why?
3. What?
4. How?
5. Usage Scenarios

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 and elsewhere
- Degree in Computer Science with focus on Natural Language Processing and Artificial Intelligence

This presentation is **purely personal** — my employer has no responsibility for any information contained here

The W3C Internationalization Tag Set (ITS) is a W3C Recommendation.

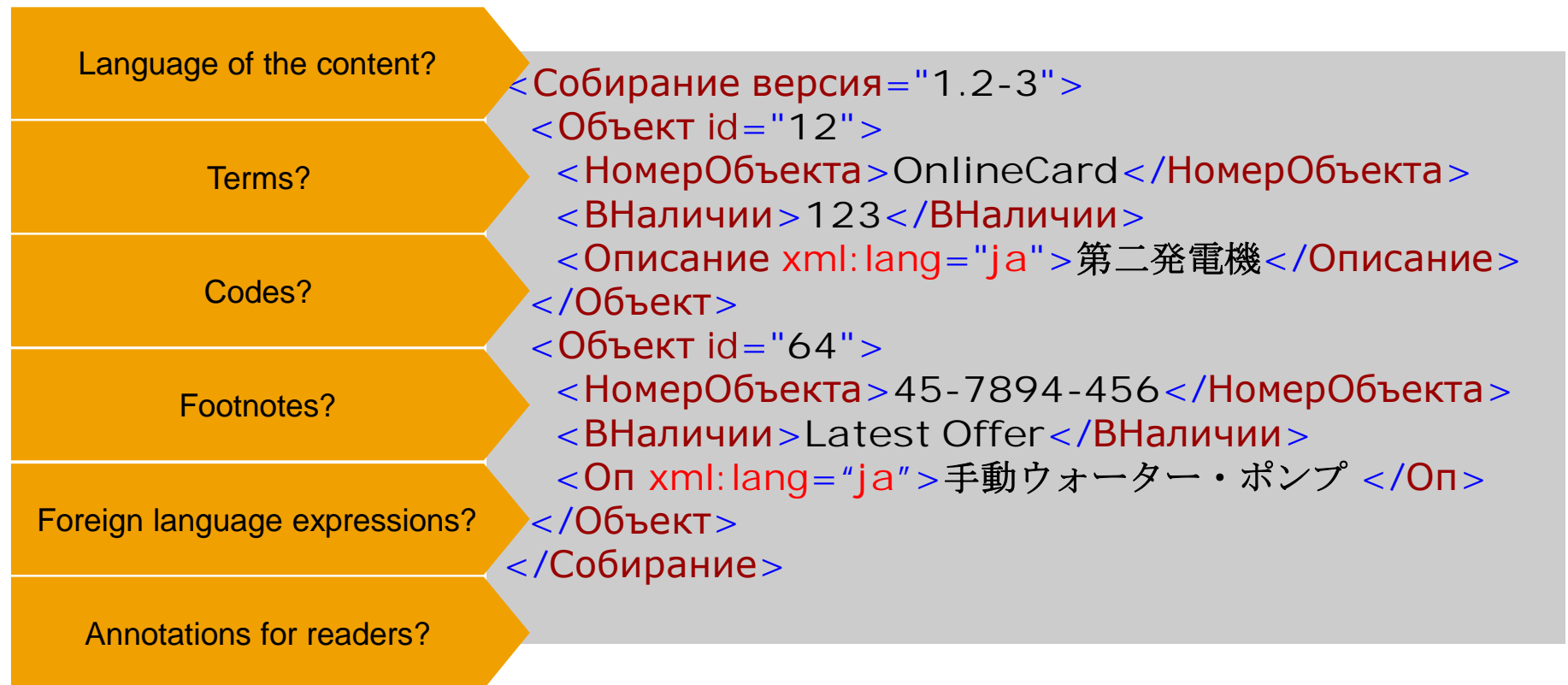
ITS helps to internationalize XML-based Contents.

Content that has been internationalized with ITS can more easily be processed by humans and machines.

ITS is an important ingredient to the W3C Note „Best Practices for XML Internationalization“

Why?

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



Adapted from Yves Savouel http://www.opentag.com/xfaq_charrep.htm#char_nonasciitag

What? (1/2)



Translate

Mark 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

What? (2/2)

The ITS data categories are valuable in themselves – that's why you do not even need to work with the ITS namespace

They cannot just be used in the ITS

They are useful also e.g. as RDF / for other non-XML data

1
Don't disturb

2
Don't reinvent the wheel

Implement different requirements with the same mechanisms

- Global / local approach for “Translate”, “Directionality”, ...

Define data categories clearly and independently

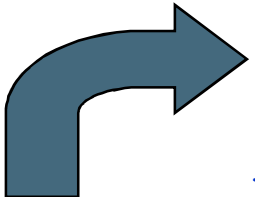
- What does “Translate” mean?
- How keep “Directionality” apart from “Ruby” etc.?

Do not enforce changes (“a secret marriage”)

- Do not disturb original content
- Add more & independent metadata

How? (2/2)

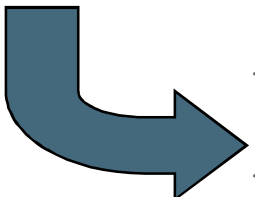
Local Approach



```
<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>
```

Best Practice Note: “Best Practices for XML Internationalization” (not covered here)

This document covers the following requirements:

- [R002 - span-like element](#), see [span](#)
- [R006 - identifying language/locale](#), see [Section 6.7: Language Information](#)
- [R007 - identifying Terms](#), see [Section 6.4: Terminology](#)
- [R008 - purpose specification/mapping](#), see [Section 5.5: Associating ITS Data Categories with Existing Markup](#)
- [R011 - bidirectional text support](#), see [Section 6.5: Directionality](#)
- [R012 - indicator of translatability](#), see [Section 6.2: Translate](#)
- [R014 - limited impact](#), see [Section 5.5: Associating ITS Data Categories with Existing Markup](#)
- [R017 - localization notes](#), see [Section 6.3: Local](#)
- [R020 - annotation markup](#), see [Section 6.6: Rub](#)
- [R025 - elements and segmentation](#), see [Section](#)

The following requirements will be addressed in [\[XML i18n BP\]](#):

- [R003 - CDATA Section](#)
- [R004 - Unique Identifier](#)
- [R005 - Handling of Entities](#)
- [R015 - Attributes and Translatable Text](#)
- [R016 - Naming Scheme](#)
- [R019 - Multilingual Documents](#)
- [R022 - Nested Elements](#)

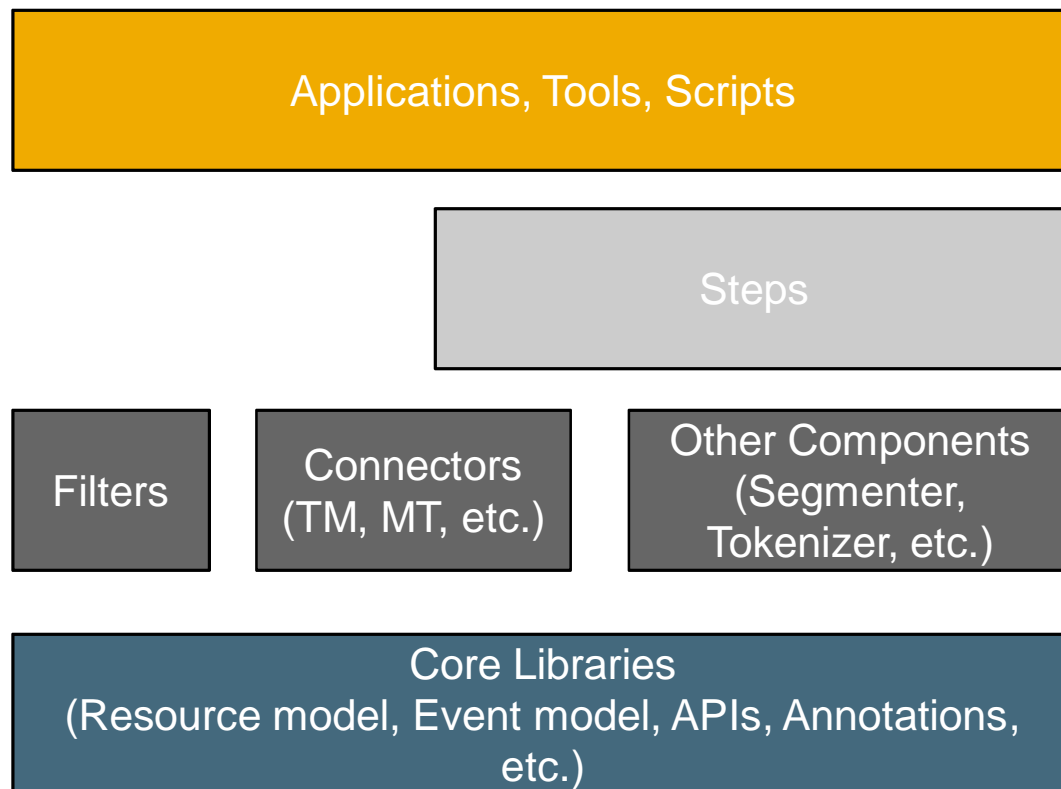
The Working Group decided not to cover the following

- [R001 - Indicator of Constraints](#)
- [R009 - Content Style](#)
- [R010 - Link to Internal/External Text](#)
- [R013 - Metrics Count](#)
- [R018 - Handling of White-Spaces](#)
- [R021 - Identifying Date and Time](#)
- [R023 - Linguistic Markup](#)
- [R024 - Variables](#)
- [R026 - Associated Objects](#)

important ones.

Usage Scenarios – Process (general)

Okapi: 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>

Usage Scenarios – Convert

ITS2XLIFF tool, see

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

converts to and from XLIFF based on ITS

Input 1): XML file with ITS “Translate” information

Output: XLIFF file

Input 2): XLIFF file with translated content

Output: original XML file with translated content

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 [Felix Sasaki](#).

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

Last Modified: 02/15/2010 21:18:26

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:

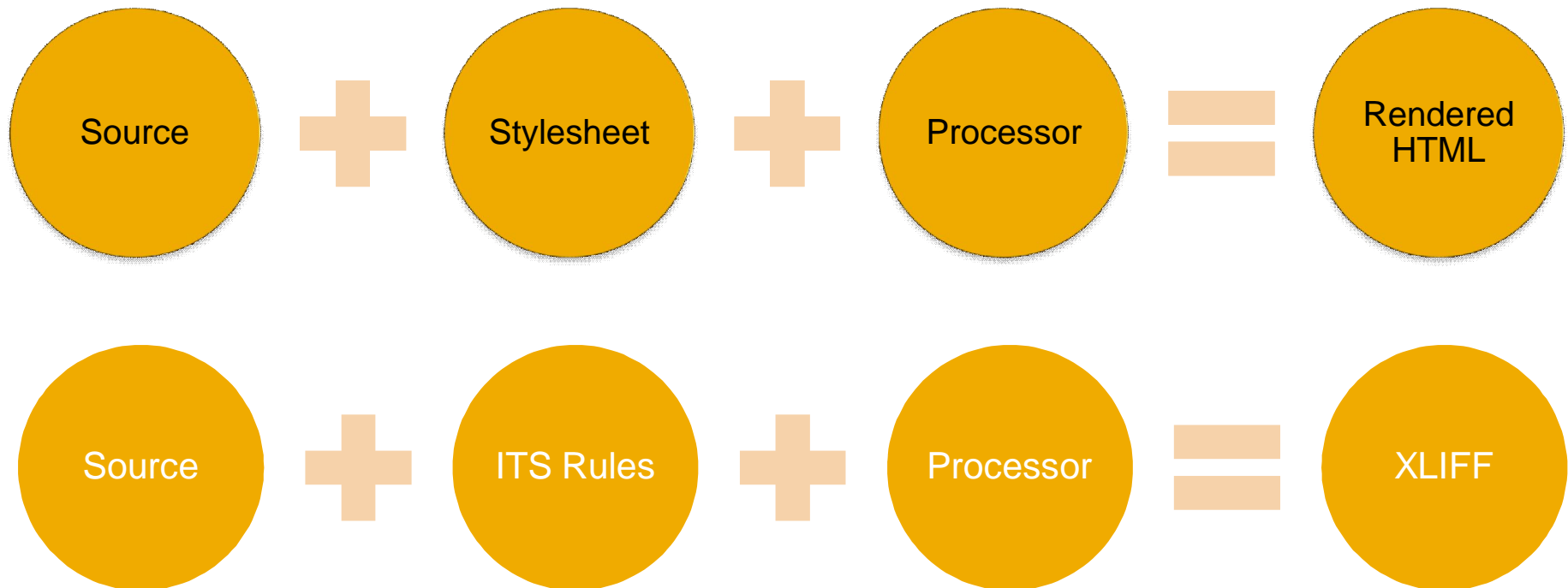
XLIFF File for conversion back to XML

Usage Scenarios – Vision (1/2)

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

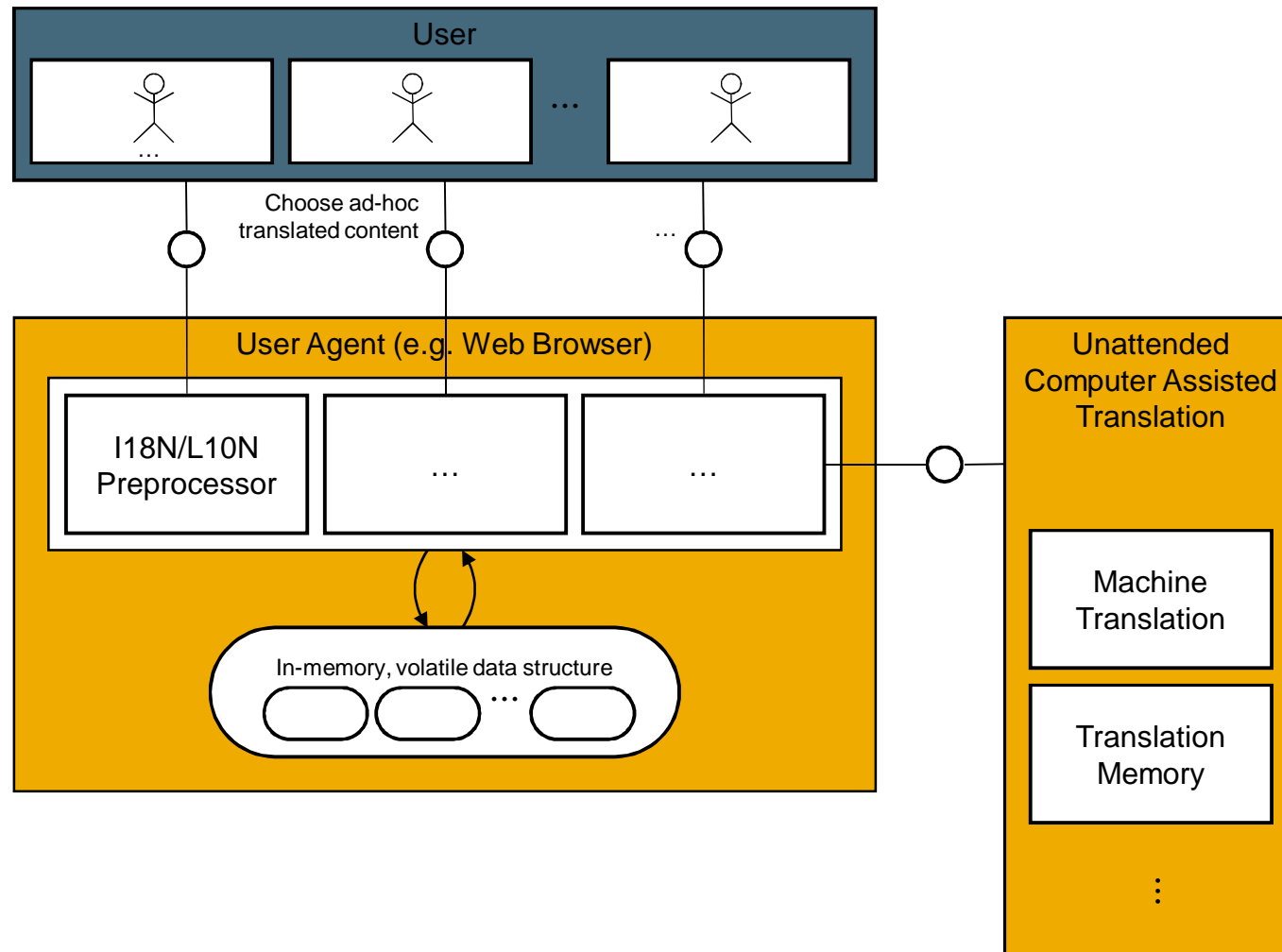
Discussion related to a MIME-type for ITS has already been started

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



Usage Scenarios – Vision (2/2)

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



Digging Deeper

Specification

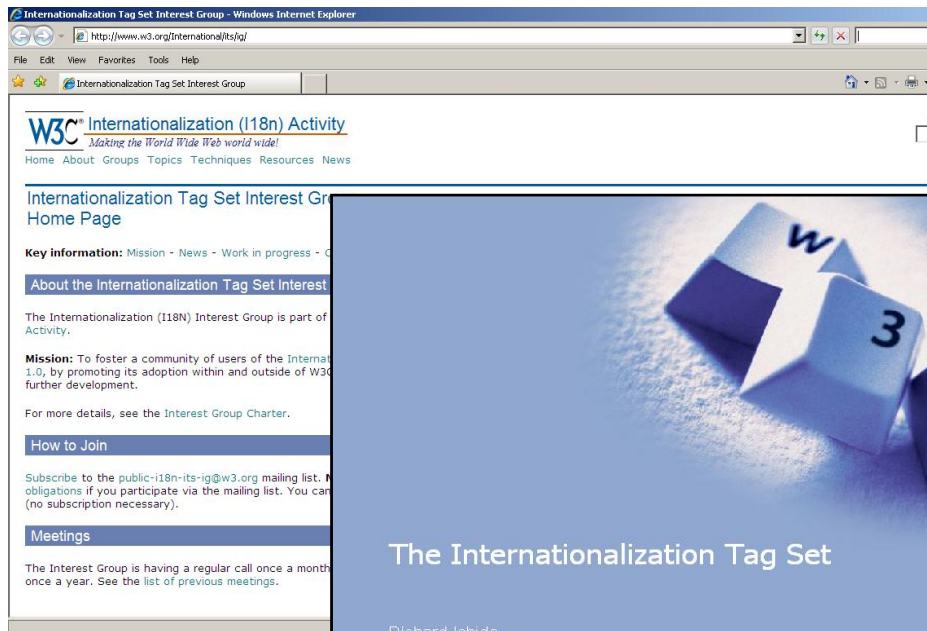
<http://www.w3.org/TR/its/>

Best Practice Note

<http://www.w3.org/TR/xml-i18n-bp>

W3C ITS Interest Group


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



The Internationalization Tag Set

Richard Ishida
W3C Internationalization Activity Lead

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



Internationalization and Localization of XML: Introducing "ITS"

Christian Lieske
Sebastian Rhatz
Felix Sasaki

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

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
http://www.tekom.de/upload/2913/LOC12_Sasaki_Lieske.pdf

Fachhochschule Potsdam
University of Applied Sciences

W3C WORLD WIDE WEB
Deutsch-Österr. Büro

THE BEST-RUN BUSINESSES RUN SAP



Thank you!



Contact

Christian Lieske

christian.lieske@sap.com
www.sap.com

Felix Sasaki

felix.sasaki@dfki.de
<http://www.dfki.de>



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.