

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

**Internationalization Tag Set 1.0 – A
New Standard for Internationalization
and Localization of XML**

Felix Sasaki
[World Wide Web Consortium](http://www.w3.org)

31st Internationalization and Unicode Conference 1 San Jose, October 2007

This presentation describes a new W3C Recommendation, the Internationalization Tag Set (ITS) 1.0, which has been developed by the W3C i18n ITS Working Group.

Internationalization Tag Set 1.0

W3C[®]

Overview

- Background: ITS 1.0 purposes / audiences
- ITS 1.0 basic "local" usage
- ITS 1.0 extended usage
- Implementations and applications

31st Internationalization and Unicode Conference

2


San Jose, October 2007

We will first give an overview of the purpose and possible audiences for ITS 1.0. Then we will describe its basic and extended usage scenarios. Finally some implementations and applications will be introduced.

Information about Internationalization and Localization in XML

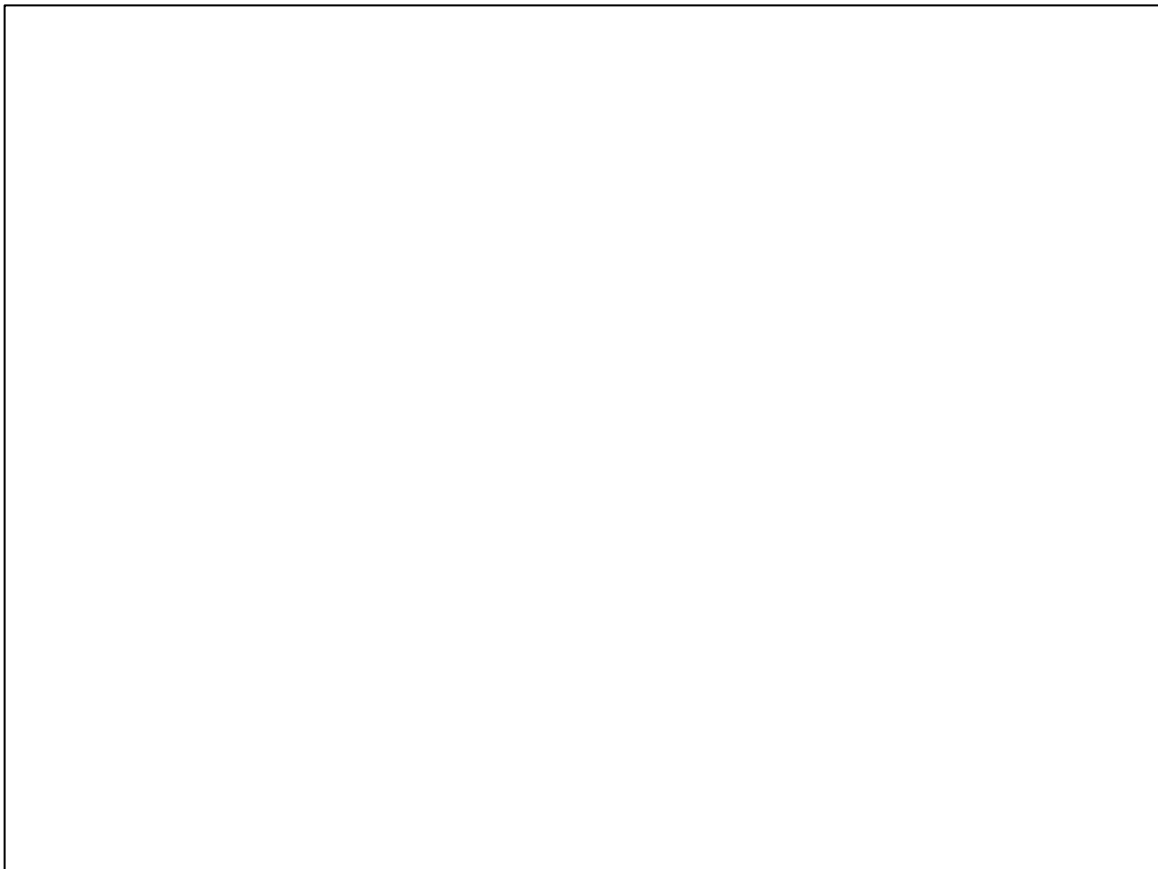
Internationalization Tag Set 1.0

Background Basic Usage Extended Usage Implementations



Background


31st Internationalization and Unicode Conference 3 San Jose, October 2007



Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Background

- ITS 1.0: "Internationalization Tag Set (ITS) Version 1.0"
<http://www.w3.org/TR/its>
- W3C Recommendation (April 2007)
produced by the W3C i18n ITS Working Group
<http://www.w3.org/International/its>


31st Internationalization and Unicode Conference4San Jose, October 2007

ITS 1.0, the "Internationalization Tag Set (ITS) Version 1.0", is a W3C Recommendation published in April 2007. It has been developed by the W3C i18n ITS Working Group.

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Background

- Target: XML Documents and Schemas
- Purpose: To express information ("data categories") for internationalization and localization of XML documents and schemas


31st Internationalization and Unicode Conference5San Jose, October 2007

ITS 1.0 targets XML documents and schemas (e.g. XML DTDs, W3C XML Schema or RELAX NG). The purpose of ITS 1.0 is to express information (in the terminology of ITS 1.0 so-called "data categories") for internationalization and localization of XML. Examples for data categories will be given later.

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Audience / Usage Scenarios

- Content authoring
- Terminology creation and translation
- Software development

31st Internationalization and Unicode Conference6San Jose, October 2007


The audience for ITS 1.0 are

- content authors who need to mark up internationalization-related or localization-related information in an XML document;
- support of terminology creation and translation in the localization process, e.g. insertion of special markers for terms; and
- software development, where the software-related material (code and / or documentation) is stored in an XML based format).

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Other topics of the Working Group

- Mostly covered by "Best Practices for XML Internationalization" document
- Draft currently under development

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

31st Internationalization and Unicode Conference7San Jose, October 2007

Besides the ITS1.0 specification covered in this talk, the ITS working group is dealing with a lot of other topics. Many of these are covered as "Best Practices for XML Internationalization" in a separate document which is currently under development. The following slide provides a list of the most important topics the working group is working on.

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

- [Scenario - Authoring Content](#)
- [Scenario - Terminology Creation and Translation](#)
- [Scenario - Software Resources](#)
- [Indicator of Constraints](#)
- [Handling entities](#)
- [Cultural aspects of the content](#)
- [Purpose specification/mapping](#)
- [Span-like elements](#)
- [Unique identifier](#)
- [Locale/language identification](#)
- [Term identification](#)
- [Indicator of translatability](#)
- [Limited impact](#)
- [CDATA section](#)
- [Links to internal/external text](#)
- [Bidirectional text support](#)
- [Indicator for metrics](#)
- [Attribute and translatable text](#)
- [Naming scheme](#)
- [Localization Notes](#)
- [Handling of white-spaces](#)
- [Multilingual Documents](#)
- [Annotation Markup](#)
- [Identifying Date and Time](#)


Complete list at

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

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

Background Basic Usage Extended Usage Implementations



Basic Usage

31st Internationalization and Unicode Conference 9 San Jose, October 2007




Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations

ITS 1.0 Data Categories



- Translate
- Localization Notes
- Terminology
- Directionality
- Ruby
- Language Information
- Elements within Text

31st Internationalization and
Unicode Conference10San Jose,
October 2007


The basic usage of ITS 1.0 is to add information, so-called "data categories" for internationalization and localization, to an XML document. The data categories on this slide are part of the ITS 1.0 specification.

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations

Sample Data category: "Translate"



- Parts of a document should **not** be translated:

```
<book>...  
<p>And he said: you need a new <quote  
  its:translate="no">motherboard</quote>  
</p> ...  
</book>
```


31st Internationalization and
Unicode Conference11San Jose,
October 2007

As an example we will introduce the "Translate" data category. It is used to express information about translatability of (parts of) an XML document. The default is that that the textual content of all elements has to be translated. The values of attributes should not be translated.

An exception to this statement can be made via an "its:translate" attribute. For example, the "its:translate" attribute with the value "no" at the <quote> element means that the content of this element should not be translated.

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Why Data Categories?

Separation of

1. prose description of ITS 1.0 categories ("data category")
2. implementation (schema language independent)
3. schema language specific declaration (XML DTDs, XML Schema, RELAX NG)

31st Internationalization and Unicode Conference12San Jose, October 2007

Why does ITS 1.0 define data categories and not markup directly? The benefit of data categories is the separation of (1) the prose description of what the ITS 1.0 information is about (the "data category"), (2) the implementation on a schema language independent level, and (3) the declaration which is specific to a schema language. The ITS 1.0 specification provides declarations for three schema language: XML DTDs, XML Schema and RELAX NG.

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

Background Basic Usage Extended Usage Implementations

W3C[®]

Example: "Translate"

1. Data category: Prose description
"Parts of a document should (not) be translated."
2. Schema language independent implementation
`<book its:translate="yes"...>... </book>`
3. Schema language specific declaration
`<!ELEMENT book ... >
<!ATTLIST book its:translate (yes|no) #IMPLIED>`


31st Internationalization and Unicode Conference 13 San Jose, October 2007

Again we give an example of the "Translate" data category. Its prose description is very simple: "Parts of a document should (not) be translated". On a schema language independent level, "Translate" is implemented via an attribute "its:translate" with the two values "yes" or "no". In the schema language XML DTDs, the attribute is declared as an optional attribute on the <book> and other, possibly all elements of a schema.

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Example: "Localization Note"

1. Data category: Prose description

"Used to provide information to localizers."
2. Schema language independent implementation

```
<book its:locNote="..."  
  locNoteType="..." ... >... </book>
```
3. Schema language specific declaration

```
<!ELEMENT book ... >  
<!ATTLIST book its:locNOTE CDATA) ...>
```


31st Internationalization and
Unicode Conference14San Jose,
October 2007

Now we give an example of the "Localization Note" data category. It is used to provide information to localizers. On a schema language independent level, "Localization Note" is implemented via the attributes "its:locNote" and "its:locNoteType". "its:locNoteType" provides the type of localization note, "description" or "alert". Declarations for these two attributes are provided In the schema language XML DTD.

Information about Internationalization and Localization in XML

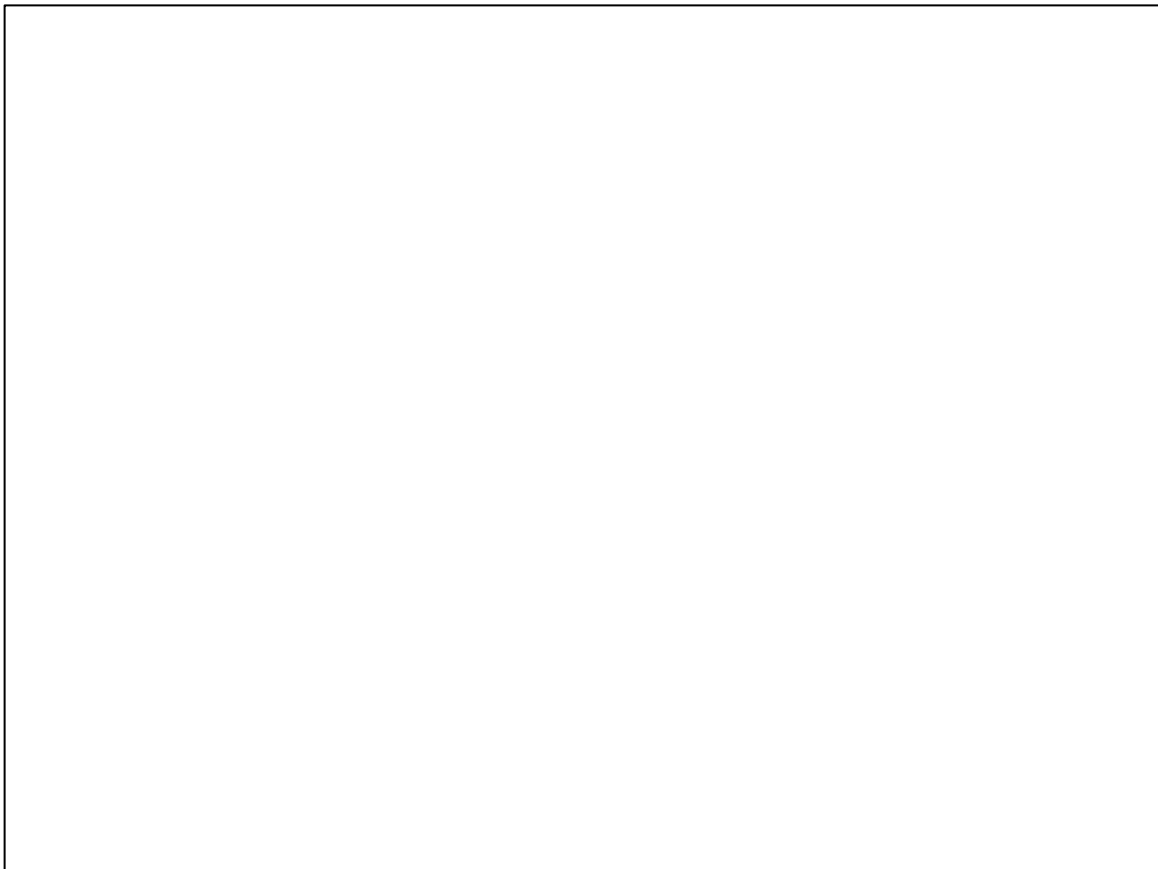
Internationalization Tag Set 1.0

Background Basic Usage **Extended Usage** Implementations




Extended Usage

31st Internationalization and Unicode Conference 15 San Jose, October 2007



Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Extended Usage of ITS

- Global usage of ITS 1.0 data categories (basic usage is called "local")
- Association between data categories and existing markup
- Pointing to existing values

31st Internationalization and Unicode Conference16San Jose, October 2007

In addition to the basic usage of ITS 1.0, there are three aspects of ITS extended usage: the global usage of ITS data categories, the association between data categories and existing markup, and pointing to existing values.

Internationalization Tag Set 1.0

Background Basic Usage **Extended Usage** Implementations

W3C[®]

Global ITS Data Categories

Using **XPath** to select parts of an XML document:

```
<someDocument ...> ...  
<its:rules ...>  
  <its:translateRule ... translate="no"  
    selector="//p[@editor='john']"/></its:rules ...>  
<!-- This rule holds for p elements which are edited by John.  
--> ...  
</someDocument>
```

Other rule elements: locNoteRule, termRule, dirRule,
...

31st Internationalization and Unicode Conference 17 San Jose, October 2007


ITS 1.0 data categories can appear in "global" positions, that means independent of a specific location in an XML document, a schema or a separate XML file. For the global usage of ITS 1.0, an `<its:rules>` element is defined. It contains data category specific child elements like `<its:translateRule>` for the "Translate" data category. The attribute "selector" at the `<its:translateRule>` Element contains an XPath expression which is used for selecting parts of an XML document. These parts are the targets to which information of the data category is applied to. In the example, the "selector attribute", in combination with the "translate='no'" attribute, expresses that the `<p>` element with the attribute "editor='john'" should not be translated.

For other data categories, ITS 1.0 defines similar rules elements. They all contain a "selector" attribute and further, data category specific markup.

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations

Combination of Global and Local Usage




```
<book ...><its:rules ...>
<its:translateRule translate="no"
selector="//quote"/></its:rules>
<body> ...
  <p>And he said: you need a new <quote
its:translate="yes">motherboard</quote>
  </p> ... </body>
</book>
```

31st Internationalization and
Unicode Conference18San Jose,
October 2007

It is possible to combine the local and global usage of ITS 1.0 data categories. In the example, the content of all `<quote>` elements should not be translated. This is expressed via `<its:translateRule>` and two attributes "translate" and "selector" at that element. As an exception, the content of the `<quote>` element should be translated. This is expressed via the "its:translate='yes'" attribute at the `<quote>` element.

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Association with Existing Markup

Association with DITA and other markup:

```
<book ...><its:rules ...>
<its:translateRule
translate="no" selector="//*[@dita:translate='no']"/>
<its:termRule term="yes" selector="//quote"/></its:rules>
<body> ...
  <p>And he said: you need a new
  <quote dita:translate="no">motherboard</quote>
  </p> ... </body>
</book>
```


31st Internationalization and
Unicode Conference19San Jose,
October 2007

The "global" usage of ITS is also a means to associate ITS data categories with existing markup. In the example, the attribute "translate='no'" at the <its:translateRule> element is associated with the "dita:translate='no'" attribute from DITA, via the "selector='//*[@dita:translate='no']'" attribute. The attribute "term='yes'" at the <its:termRule> element implements the "Terminology" data category. It is associated with all <quote> elements via the "selector='//quote'" attribute. This means that all <quote> elements are interpreted as terms.

DITA ("Darwin Information Typing Architecture") is a more and more widely used XML vocabulary especially designed for technical documentation. DITA documents are a primary target for localization, and DITA provides relevant markup like the "dita:translate" attribute.

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Pointing to Existing Values

Pointing to existing localization notes:

```
<myDoc ...><its:rules ...>
<its:locNoteRule
locNoteType="alert" selector="//*[@loc-n]"
locNotePointer="@loc-n"/>...</its:rules>
...
<span loc-n="...">...</span>
...
</myDoc>
```

31st Internationalization and
Unicode Conference20San Jose,
October 2007


Finally, it is possible to point to existing values in a document. An example is given for the "Localization Note" data category. The "locNotePointer" attribute at the <its:locNoteRule> element is used to point to the value of the "loc-n" attribute.

Association with markup and pointing to values make it possible to apply ITS without an impact on existing XML documents, and re-use markup (like the DITA "dita:translate" attribute) from existing XML vocabularies.

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Implementations and Applications


31st Internationalization and Unicode Conference21San Jose, October 2007

Now we will present some implementations and applications of ITS 1.0. The data on the next slide is taken from the ITS 1.0 test suite, see <http://www.w3.org/International/its/tests/> .

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



ITS Test Suite

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


- Translate ([input](#), [output](#))
- Localization Note ([input](#), [output](#))
- Terminology ([input](#), [ouput](#))
- Language Information ([input](#), [output](#))
- Elements within Text ([input](#), [output](#))

31st Internationalization and
Unicode Conference22San Jose,
October 2007

There is sample input for five data categories: "Translate", Localization Note, Terminology, Language Information and Elements within Text. The output format is specific to the ITS 1.0 test suite. It makes explicit what information is available after ITS 1.0 processing.

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations



Implementations

- XSLT implementations for all data categories
 - [Spritser](#)
 - [ITSImpl](#)
- [Rainbow XML Filter](#)
 - preparation for translation
 - XLIFF extraction
 - spell-checking
 - etc.

31st Internationalization and Unicode Conference23San Jose, October 2007


During the development of the ITS 1.0 specification, four implementations have been created. There are two XSLT-based implementations, "Spritser" and "ITSImpl", which cover all ITS 1.0 data categories. The "Rainbow XML Filter" covers the "Translate", Localization Note, Terminology and Elements within Text data categories. This filter has many application scenarios which are listed on this slide.

Information about Internationalization and Localization in XML

Internationalization Tag Set 1.0

BackgroundBasic UsageExtended UsageImplementations

Implementations



- [XTM](#)
 - Web-based translation tool system
- ITS2Xliff (under development!)
- Further information at <http://www.w3.org/International/its/links>
also about:
- **Your implementation of ITS 1.0!**
 - [Please let us know](#)

31st Internationalization and Unicode Conference24San Jose, October 2007

XTM covers the "Translate" and "Elements within Text" data categories.

ITS2Xliff is a new implementation under development. Its purpose is the generation of XLIFF ("XML Localization Interchange File Format") documents from XML files with local or global ITS markup.

Further information about implementations and applications is available at the link on this slide. We hope that sooner or later we will also be able to add a link to your implementation of ITS 1.0!

Information about Internationalization and Localization in XML

