**1 Introduction**

*This section is informative.*

The Internationalization Tag Set (ITS) is a technology designed support the growth of the Multilingual Web by improving the ease and efficiency with which web content is localised, i.e. adapted for the language, cultural and other requirements of a specific target audience (known as a ‘locale’). ITS achieves this by improving interoperability between systems commonly involved in activities related to localisation. These activities include, but are not limited to: internationalisation, where the original content, known as source content, is adapted to ease its localisation; translation, where the source content is translated by human and/or automated means into one or more target locale languages; and the processing of the translated content to assess its quality and prepare it for publication and for reuse as a language resource in future localization activities.

erange of activities supported by the ITS specification derives from the requirements and use casesidenitfied in

The ITS 2.0 specification defines a set of concepts (termed “ITS data categories”) that support different functions related to localisation activities. It also defines the implementation of these concepts as a set of elements and attributes, termed the Internationalization Tag Set (ITS), that can be used as markup with any HTML5 and XML documents that may be subject to a localisation activity.

The document therefore provides implementations for ITS as a mark-up schema for use with HTML5, and as schema specified inthe languages XML DTD [[XML 1.0]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xml10spec), XML Schema [[XML Schema]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xmlschema1) and RELAX NG [[RELAX NG]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#relaxng) for use with XML.

The Semantic Web and the open provision of linguistic knowledge as Linked Data may support many localisation activities so the ITS specification offers a serialisation into RDF and will support mappings into schema supporting linguistic knowledge such as NIF.

**[[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "contents)1.1 Relation to ITS 1.0 and New Principles**

ITS 2.0 is the successor to ITS1.0 [].**[[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "contents)1.1.1 Relation to ITS 1.0**

ITS 2.0 has the following relations to ITS 1.0:

* It adopts and maintains the following principles from ITS 1.0:
  + It adopts the use of data categories to define discrete units of functionality
  + It adopts the separation of data category definition from the mapping of the data category to a given content format
  + It adopts the conformance principle of ITS1.0 that an implementation only needs to implement one data category to claim conformance to ITS 2.0
* ITS 2.0 supports all ITS 1.0 data category definitions..
* ITS 2.0 adds a number of new data categories not found in ITS 1.0.
* While ITS 1.0 addressed only XML, ITS 2.0 specifies implementations of data categories in *both* XML *and* HTML5.
* Where ITS 1.0 data categories are implemented in XML, the implementation must be conformant with the ITS 1.0 approach to XML to claim conformance to ITS 2.0.

**[[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "contents)1.1.2 New Principles**

ITS 2.0 also adds the following principles and features not found in ITS 1.0:

* ITS 2.0 data categories are intended to be format neutral, with support for XML, HTML5, RDFa, and NIF: a data category implementation only needs to support a single content format mapping in order to support a claim of ITS 2.0 conformance.
* ITS 2.0 provides algorithms to generate RDFa and NIF out of HTML5 or XML with ITS 2.0 metadata.
* A global implementation of ITS 2.0 requires at least the XPath version 1.0. Other versions of XPath or other query languages (e.g., CSS selectors) can be expressed via a dedicated [queryLanguage](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "queryLanguage) attribute.
* Additional flexibility in the management of ITS rules through the introduction of rule parameters.

As of the time of this writing, the new data categories included in ITS 2.0 are:

* [Domain](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#domain)
* [Disambiguation](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#Disambiguation)
* [Locale Filter](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#LocaleFilter)
* [Provenance](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#Provenance)
* [Text Analysis Annotation](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#TextAnalyisAnnotation)

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**1.2 Motivation for ITS**

Content or software that is authored in one language (the source language) is often made available in additional languages or adapted with regard to other cultural aspects. This is done through a process called localization, where the original material is translated and adapted to the target audience.

In addition, document formats expressed by schemas may be used by people in different parts of the world, and these people may need special markup to support the local language or script. For example, people authoring in languages such as Arabic, Hebrew, Persian, or Urdu need special markup to specify directionality in mixed direction text.

From the viewpoints of feasibility, cost, and efficiency, it is important that the original material should be suitable for localization. This is achieved by appropriate design and development, and the corresponding process is referred to as internationalization. For a detailed explanation of the terms “localization” and “internationalization”, see [[l10n i18n]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "geo-i18n-l10n" \o ").

[Ed. note: Note: This should refer to the best practice document as well, when ready.]

The increasing usage of XML as a medium for documentation-related content (e.g. DocBook and DITA as formats for writing structured documentation, well suited to computer hardware and software manuals) and software-related content (e.g. the eXtensible User Interface Language [[XUL]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xul)) creates challenges and opportunities in the domain of XML internationalization and localization.

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**1.2.1 Typical Problems**

The following examples sketch one of the issues that currently hinder efficient XML-related localization: the lack of a standard, declarative mechanism that identifies which parts of an XML document need to be translated. Tools often cannot automatically perform this identification.

Example 1: Document with partially translatable content

In this document it is difficult to distinguish between those string elements that are translatable and those that are not. Only the addition of an explicit flag could resolve the issue.

<resources>

 <section id="Homepage">

  <arguments>

   <string>page</string>

   <string>childlist</string>

  </arguments>

  <variables>

   <string>POLICY</string>

   <string>Corporate Policy</string>

  </variables>

  <keyvalue\_pairs>

   <string>Page</string>

   <string>ABC Corporation - Policy Repository</string>

   <string>Footer\_Last</string>

   <string>Pages</string>

   <string>bgColor</string>

   <string>NavajoWhite</string>

   <string>title</string>

   <string>List of Available Policies</string>

  </keyvalue\_pairs>

 </section>

</resources>

[Source file: [examples/xml/EX-motivation-its-1.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-motivation-its-1.xml)]

Example 2: Document with partially translatable content

Even when metadata are available to identify non-translatable text, the conditions may be quite complex and not directly indicated with a simple flag. Here, for instance, only the text in the nodes matching the expression //component[@type!='image']/data[@type='text'] is translatable.

<dialogue xml:lang="en-gb">

 <rsrc id="123">

  <component id="456" type="image">

   <data type="text">images/cancel.gif</data>

   <data type="coordinates">12,20,50,14</data>

  </component>

  <component id="789" type="caption">

   <data type="text">Cancel</data>

   <data type="coordinates">12,34,50,14</data>

  </component>

  <component id="792" type="string">

   <data type="text">Number of files: </data>

  </component>

 </rsrc>

</dialogue>

[Source file: [examples/xml/EX-motivation-its-2.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-motivation-its-2.xml)]

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**1.3 Users and Usages of ITS**

**[[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "contents)1.3.1 Potential Users of ITS**

The ITS specification aims to provide different types of users with information about what markup should be supported to enable worldwide use and effective internationalization and localization of content. The following paragraphs sketch these different types of users, and their usage of ITS. In order to support all of these users, the information about what markup should be supported to enable worldwide use and effective localization of content is provided in this specification in two ways:

* abstractly in the data category descriptions: [Section 6: Description of Data Categories](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "datacategory-description)
* concretely in the ITS schemas: [Appendix E: Schemas for ITS](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#its-schemas)

**1.3.1.1Schema developers starting a schema from the ground up**

This type of user will find proposals for attribute and element names to be included in their new schema (also called "host vocabulary"). Using the attribute and element names proposed in the ITS specification may be helpful because it leads to easier recognition of the concepts represented by both schema users and processors. It is perfectly possible, however, for a schema developer to develop his own set of attribute and element names. The specification sets out, first and foremost, to ensure that the required markup is available, and that the behavior of that markup meets established needs.

**1.3.1.2Schema developers working with an existing schema**

This type of user will be working with schemas such as DocBook, DITA, or perhaps a proprietary schema. The ITS Working Group has sought input from experts developing widely used formats such as the ones mentioned.

**Note:**

The question "How to use ITS with existing popular markup schemes?" is covered in more details (including examples) in a separate document: [[XML i18n BP]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "xml-i18n-bp" \o "Best Practices for XML Internationalization).

Developers working on existing schemas should check whether their schemas support the markup proposed in this specification, and, where appropriate, add the markup proposed here to their schema.

In some cases, an existing schema may already contain markup equivalent to that recommended in ITS. In this case it is not necessary to add duplicate markup since ITS provides mechanisms for associating ITS markup with markup in the host vocabulary which serves a similar purpose (see [Section 5.6: Associating ITS Data Categories with Existing Markup](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#associating-its-with-existing-markup)). The developer should, however, check that the behavior associated with the markup in their own schema is fully compatible with the expectations described in this specification.

**1.3.1.3Vendors of content-related tools**

This type of user includes companies which provide tools for authoring, translation or other flavors of content-related software solutions. It is important to ensure that such tools enable worldwide use and effective localization of content. For example, translation tools should prevent content marked up as not for translation from being changed or translated. It is hoped that the ITS specification will make the job of vendors easier by standardizing the format and processing expectations of certain relevant markup items, and allowing them to more effectively identify how content should be handled.

**1.3.1.4Content producers**

This type of user comprises authors, translators and other types of content author. The markup proposed in this specification may be used by them to mark up specific bits of content. Aside: The burden of inserting markup can be removed from content producers by relating the ITS information to relevant bits of content in a global manner (see [global, rule-based approach](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "selection-global)). This global work, however, may fall to information architects, rather than the content producers themselves.

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**1.3.2 Ways to Use ITS**

The ITS specification proposes several mechanisms for supporting worldwide use and effective internationalization and localization of content. We will sketch them below by looking at them from the perspectives of certain user types. For the purpose of illustration, we will demonstrate how ITS can indicate that certain parts of content should or should not be translated.

* A content author uses an attribute on a particular element to say that the text in the element should not be translated.

Example 3: Use of ITS by content author

The its:translate="no" attributes indicate that the path and the cmd elements should not be translated.

<help

xmlns:its="http://www.w3.org/2005/11/its"

  its:version="2.0">

 <head>

  <title>Building the Zebulon Toolkit</title>

 </head>

 <body>

  <p>To re-compile all the modules of the Zebulon toolkit you need to go in the

   <path

     its:translate="no">\Zebulon\Current Source\binary</path> directory.

   Then from there, run batch file <cmd

     its:translate="no">Build.bat</cmd>.</p>

 </body>

</help>

[Source file: [examples/xml/EX-ways-to-use-its-1.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-ways-to-use-its-1.xml)]

* A content author or information architect uses markup at the top of the document to identify a particular type of element or context in which the content should not be translated.

Example 4: Use of ITS by information architect

The [translateRule](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "translateRule) element is used in the header of the document to indicate that none of the path or cmd elements should be translated.

<help

xmlns:its="http://www.w3.org/2005/11/its"

  its:version="2.0">

 <head>

  <title>Building the Zebulon Toolkit</title>

  <its:rules version="2.0">

   <its:translateRule selector="//path | //cmd" translate="no"/>

  </its:rules>

 </head>

 <body>

  <p>To re-compile all the modules of the Zebulon toolkit you need to go in the

   <path>\Zebulon\Current Source\binary</path> directory.

   Then from there, run batch file <cmd>Build.bat</cmd>.</p>

 </body>

</help>

[Source file: [examples/xml/EX-ways-to-use-its-2.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-ways-to-use-its-2.xml)]

* A processor may insert markup at the top of the document which links to ITS information outside of the document.

Example 5: Use of ITS by processor

A [rules](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "rules) element is inserted in the header of the document. It has a XLink [href](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#rules.attributes) attribute used to link to an [ITS external rule](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#link-external-rules) document.

<help

xmlns:its="http://www.w3.org/2005/11/its"

xmlns:xlink="http://www.w3.org/1999/xlink"

  its:version="2.0">

 <head>

  <title>Building the Zebulon Toolkit</title>

  <its:rules version="2.0" xlink:href="EX-ways-to-use-its-4.xml"/>

 </head>

 <body>

  <p>To re-compile all the modules of the Zebulon toolkit you need to go in the

   <path>\Zebulon\Current Source\binary</path> directory.

   Then from there, run batch file <cmd>Build.bat</cmd>.</p>

 </body>

</help>

[Source file: [examples/xml/EX-ways-to-use-its-3.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-ways-to-use-its-3.xml)]

Example 6: ITS rule file shared by different documents

The [rules](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "rules) element contains several ITS rules that are common to different documents. One of them is a [translateRule](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#translateRule) element that indicates that no path or cmd element should be translated.

<its:rules

xmlns:its="http://www.w3.org/2005/11/its"  version="2.0">

 <its:translateRule selector="//path | //cmd" translate="no"/>

</its:rules>

[Source file: [examples/xml/EX-ways-to-use-its-4.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-ways-to-use-its-4.xml)]

* A schema developer integrates ITS markup declarations in his schema to allow users to indicate that specific parts of the content should not be translated.

Example 7: An XSD schema with ITS declaration

The declarations for the [translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "att.local.no-ns.attribute.translate) attribute is added to a group of common attributes commonAtts. This allows to use the [translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.local.no-ns.attribute.translate) attribute within the documents like in [Example 3](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#EX-ways-to-use-its-1).

<xs:schema

xmlns:its="http://www.w3.org/2005/11/its"

xmlns:xs="http://www.w3.org/2001/XMLSchema"  elementFormDefault="qualified">

 <xs:import namespace="http://www.w3.org/2005/11/its" schemaLocation="its.xsd"/>

 <xs:attributeGroup name="commonAtts">

  <xs:attributeGroup ref="its:att.local.with-ns.attribute.translate"/>

  <xs:attribute name="id" type="xs:ID" use="optional"/>

 </xs:attributeGroup>

 <xs:element name="help">

  <xs:complexType>

   <xs:sequence>

    <xs:element name="head">

     <xs:complexType>

      <xs:sequence>

       <xs:element name="title" type="xs:string"/>

      </xs:sequence>

      <xs:attributeGroup ref="commonAtts"/>

     </xs:complexType>

    </xs:element>

    <xs:element name="body">

     <xs:complexType>

      <xs:choice minOccurs="1" maxOccurs="unbounded">

       <xs:element name="p">

        <xs:complexType mixed="true">

         <xs:choice minOccurs="0" maxOccurs="unbounded">

          <xs:element ref="path"/>

          <xs:element ref="cmd"/>

         </xs:choice>

         <xs:attributeGroup ref="commonAtts"/>

        </xs:complexType>

       </xs:element>

      </xs:choice>

     </xs:complexType>

    </xs:element>

   </xs:sequence>

   <xs:attributeGroup ref="its:att.version.attribute.version"/>

  </xs:complexType>

 </xs:element>

 <xs:element name="path">

  <xs:complexType mixed="true">

   <xs:attributeGroup ref="commonAtts"/>

  </xs:complexType>

 </xs:element>

 <xs:element name="cmd">

  <xs:complexType mixed="true">

   <xs:attributeGroup ref="commonAtts"/>

  </xs:complexType>

 </xs:element>

</xs:schema>

[Source file: [examples/xml/EX-ways-to-use-its-5.xsd](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-ways-to-use-its-5.xsd)]

The first two approaches above can be likened to the use of CSS in [[XHTML 1.0]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xhtml10). Using a style attribute, an XHTML content author may assign a color to a particular paragraph. That author could also have used the style element at the top of the page to say that all paragraphs of a particular class or in a particular context would be colored red.

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**1.4 Usage in HTML5**

ITS 2.0 adds support for usage in HTML5. In HTML5, ITS local selection is realized via dedicated, [data category specific attributes](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "att.local.html5).

For the so-called “[global approach](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#basic-concepts-selection-global)” in HTML5, this specification defines a link type for referring to files with global rules. These rules are then processed as described in [Section 5.2.2: Global selection within HTML5](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selection-global-html5).

Example 8: Using ITS global rules in HTML5

The link element points to the rules file EX-translateRule-html5-1.xml The rel attribute identifies the ITS specific link relation its-rules.

<!DOCTYPE html>

<html lang="en">

 <head>

  <meta charset="utf-8"><meta>

  <title>Translate flag global rules example</title>

  <link href="EX-translateRule-html5-1.xml" rel="its-rules"><link>

 </head>

 <body>

  <p>This sentence should be translated, but code names like the <code>span</code> element should not be translated.</p>

 </body>

</html>

[Source file: [examples/html5/EX-translate-html5-global-1.html](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/html5/EX-translate-html5-global-1.html)]

Example 9: ITS rules file linked from HTML5

The rules file linked in [Example 8](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "EX-translate-html5-global-1).

<its:rules

xmlns:its="http://www.w3.org/2005/11/its"  version="2.0">

 <its:translateRule translate="no" selector="//h:code"/>

</its:rules>

[Source file: [examples/html5/EX-translateRule-html5-1.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/html5/EX-translateRule-html5-1.xml)]

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**1.4.1 Support for legacy HTML content**

ITS 2.0 does not define how to use ITS in HTML versions prior version 5. Users are encouraged to migrate their content to HTML5 or XHTML. While it is possible to use its-\* attributes introduced for HTML5 in older versions of HTML (such as 3.2 or 4.01) and pages using these attributes will work without any problems, its-\* attributes will be marked as invalid in validators.

**[[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "contents)1.5 Out of Scope**

The definition of what a localization process or localization parameters must address is outside the scope of this standard and it does not address all of the mechanisms or data formats (sometimes called Localization Properties) that may be needed to configure localization workflows or process specific formats. However, it does define standard data categories that may be used in defining localization workflows or processing specific formats.

**Note:**

“XML localization properties” is a generic term to name the mechanisms and data formats that allow localization tools to be configured in order to process a specific XML format. Examples of XML localization properties are the Trados “DTD Settings” file, and the SDLX “Analysis” file.

**[[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "contents)1.6 Important Design Principles**

Abstraction via *data categories*: ITS defines data categories as an abstract notion for information needed for the internationalization and localization of XML schemas and documents and HTML5 documents. This abstraction is helpful in realizing independence from any one particular implementation (e.g., as an element or attribute). (See [Section 3.3: Data category](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "def-datacat) for a definition of the term data categories, [Section 6: Description of Data Categories](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#datacategory-description) for the definition of the various ITS data categories, and subsections in [Section 6: Description of Data Categories](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#datacategory-description) for the data category implementations.)

Powerful *selection mechanism:* For ITS markup that appears in an XML instance, which XML nodes the ITS-related information pertains to must be clearly defined. Thus, ITS defines [selection](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#termdef-selection) mechanisms to specify to what parts of an XML document an ITS data category and its values should be applied. Selection relies on the information which is given in the XML Information Set [[XML Infoset]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xmlinfoset). ITS applications may implement inclusion mechanisms such as XInclude or DITA's [[DITA 1.0]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#dita10) conref.

Content authors, for example, need a simple way to work with the [Translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#trans-datacat) data category in order to express whether the content of an element or attribute should be translated or not. Localization managers, on the other hand, need an efficient way to manage translations of large document sets based on the same schema. These needs could by realized by a specification of defaults for the [Translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#trans-datacat) data category along with exceptions to those defaults (e.g. all p elements should be translated, but not p elements inside of an index element).

To meet these requirements this specification introduces mechanisms that add ITS information to XML documents, see [Section 5: Processing of ITS information](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#its-processing). These mechanisms also provide a means for specifying ITS information for attributes (a task for which no standard means previously existed).

The ITS selection mechanisms allows you to provide information about content [locally](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selection-local) (specified at the XML or HTML element to which it pertains) or [globally](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selection-global) (specified in another part of the document). Global selection mechanisms can be in the same document, or in a separate file.

*No dedicated extensibility*: It may be useful or necessary to extend the set of information available for internationalization or localization purposes beyond what is provided by ITS. This specification does not define a dedicated extension mechanism, since ordinary XML mechanisms (e.g. XML Namespaces [[XML Names]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xmlns)) may be used.

*Ease of integration*:

* ITS follows the example from [section 4](http://www.w3.org/TR/2001/REC-xlink-20010627/#att-method) of [[XLink 1.1]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xlink1), by providing mostly global attributes for the implementation of ITS data categories. Avoiding elements for ITS purposes as much as possible ensures ease of integration into existing markup schemes, see [section 3.14](http://www.w3.org/TR/itsreq/#impact) in [[ITS REQ]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#itsreq). Only for some requirements do additional child elements have to be used, see for example [Section 6.6: Ruby](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#ruby-annotation).
* ITS has no dependency on technologies which are still under development.
* ITS fits with existing work in the W3C architecture (e.g. use of [[XPath 1.0]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xpath) for the selection mechanism).

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**1.7 Development of this Specification**

This specification has been developed using the ODD (*One Document Does it all*) language of the Text Encoding Initiative ( [[TEI]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "tei" \o "Text Encoding Initiative Guidelines development version (P5))). This is a literate programming language for writing XML schemas, with three characteristics:

1. The element and attribute set is specified using an XML vocabulary which includes support for macros (like DTD entities, or schema patterns), a hierarchical class system for attributes and elements, and creation of modules.
2. The content models for elements and attributes are written using embedded RELAX NG XML notation.
3. Documentation for elements, attributes, value lists etc. is written inline, along with examples and other supporting material.

XSLT transformations are provided by the TEI to create documentation into HTML, XSL FO or LaTeX forms, and to generate RELAX NG documents and DTD. From the RELAX NG documents, James Clark's [trang](http://www.thaiopensource.com/relaxng/trang.html) can be used to create XML Schema documents.

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**2 Basic Concepts**

*This section is informative.*

**[[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "contents)2.1 Selection**

Information (e.g. "translate this") captured by ITS markup (e.g. its:translate='yes') always pertains to one or more XML or HTML nodes (primarily element and attribute nodes). In a sense, ITS markup “selects” the relevant node(s). Selection may be explicit or implicit. ITS distinguishes two approaches to selection: (1) local, and (2) using global rules.

The mechanisms defined for ITS selection resemble those defined in [[CSS 2.1]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "css2-1" \o "Cascading Style Sheets, level 2 revision 1 CSS 2.1 Specification). The local approach can be compared to the style attribute in HTML/XHTML, and the approach with global rules is similar to the style element in HTML/XHTML. ITS usually uses XPath for identifying nodes although CSS and other query languages can be used if supported by application. Thus,

* the local approach puts ITS markup in the relevant element of the host vocabulary (e.g. the author element in DocBook)
* the [rule-based, global approach](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#selection-global) puts the ITS markup in elements defined by ITS itself (namely the [rules](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#rules) element)

ITS markup can be used with XML documents (e.g. a DocBook article), or schemas (e.g. an XML Schema document for a proprietary document format). Since each usage defines some specific requirements, ITS markup may take different shapes.

[Ed. note: In the paragraph above, we need an explanation of what “different shapes” means.]

The following two examples sketch the distinction between the local and global approaches, using the [translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.local.no-ns.attribute.translate) as one example of ITS markup.

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**2.1.1 Local Approach**

The document in [Example 10](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "EX-basic-concepts-1) shows how a content author may use the ITS [translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.local.no-ns.attribute.translate) attribute to indicate that all content inside the author element should be protected from translation. Translation tools that are aware of the meaning of this attribute can then screen the relevant content from the translation process.

Example 10: ITS markup on elements in an XML document (local approach)

<article

xmlns="http://docbook.org/ns/docbook"

xmlns:its="http://www.w3.org/2005/11/its"

  its:version="2.0" version="5.0" xml:lang="en">

 <info>

  <title>An example article</title>

  <author

    its:translate="no">

   <personname>

    <firstname>John</firstname>

    <surname>Doe</surname>

   </personname>

   <affiliation>

    <address>

     <email>foo@example.com</email>

    </address>

   </affiliation>

  </author>

 </info>

 <para>This is a short article.</para>

</article>

[Source file: [examples/xml/EX-basic-concepts-1.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-basic-concepts-1.xml)]

For this example to work, the schema developer will need to add the [translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.local.no-ns.attribute.translate) attribute to the schema as a common attribute or on all the relevant element definitions. Note how there is an expectation in this case that inheritance plays a part in identifying which content does have to be translated and which does not. Tools that process this content for translation will need to implement the expected inheritance.

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**2.1.2 Global Approach**

The document in [Example 11](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "EX-basic-concepts-2) shows a different approach to identifying non-translatable content, similar to that used with a style element in [[XHTML 1.0]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xhtml10), but using an ITS-defined element called [rules](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#rules). It works as follows: A document can contain a [rules](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#rules) element (placed where it does not impact the structure of the document, e.g., in a “head” section). It contains one or more ITS rule elements (for example [translateRule](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#translateRule)). Each of these specific elements contains a [selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.selector.attribute.selector) attribute. As its name suggests, this attribute selects the node or nodes to which a corresponding ITS information pertains. The values of ITS selector attributes are XPath absolute location paths (or CSS selectors if [queryLanguage](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#queryLanguage) is set to "css"). Information for the handling of namespaces in these path expressions is taken from namespace declarations [[XML Names]](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#xmlns) at the current rule element.

[Ed. note: The following needs to be updated to allow for the use of queryLanguage, since we can use CSS selectors now. Perhaps the correction is as simple as changing the first sentence to read “When using XPath values for ITS selector attributes (see queryLanguage for a discussion of alternate values), note that the values are XPath absolute location paths…” Would that do? Also, might we move this to the end of the section?]

**Note:**

Caveat Related to XSLT-based Processing of ITS Selector Attributes

The values of ITS [selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.selector.attribute.selector) attributes are XPath absolute location paths. Accordingly, the following is a legitimate value:

myElement/descendant-or-self::\*/@\*

Unfortunately, values like this cause trouble when they are used in XSLT-based processing of ITS where the values of the ITS [selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.selector.attribute.selector) attributes are used as values of match attributes of XSLT templates. The reason for this is the following: match attributes may only contain a restriction/subset of XPath expressions, so-called [patterns](http://www.w3.org/TR/xslt#patterns).

Basically the following restrictions hold for patterns:

* only axes "child" or "attribute" allowed
* "//" or "/" possible
* id() or key() function possible
* predicates possible

Using only XSLT patterns in ITS [selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.selector.attribute.selector) attributes helps to avoid this issue. In many cases, this is possible by using patterns with predicates. The value above may for example be rewritten as follows:

\*[self::myElement]/@\* | myElement//\*/@\*

Example 11: ITS global markup in an XML document (rule-based approach)

<myTopic

xmlns:its="http://www.w3.org/2005/11/its"

xmlns="myNamescapeURI" id="topic01" xml:lang="en-us">

 <prolog>

  <title>Using ITS</title>

  <its:rules version="2.0">

   <its:translateRule selector="//n:term" translate="no"/>

  </its:rules>

 </prolog>

 <body>

  <p>ITS defines <term>data category</term> as an abstract

  concept for a particular type of information for

  internationalization and localization of XML schemas and

  documents.</p>

 </body>

</myTopic>

[Source file: [examples/xml/EX-basic-concepts-2.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-basic-concepts-2.xml)]

For this approach to work, the schema developer needs to add the [rules](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#rules) element and associated markup to the schema. In some cases global rules may be sufficient to allow the schema developer to avoid adding other ITS markup (such as an [translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.local.no-ns.attribute.translate) attribute) to the elements and attributes in the schema. However, it is likely that authors will want to use attributes on markup from time to time to override the general rule.

For specification of the [Translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#trans-datacat) data category information, the contents of the [rules](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#rules) element would normally be designed by an information architect familiar with the document format and familiar with, or working with someone familiar with, the needs of the localization group.

The global, rule-based approach has the following benefits:

* Content authors do not have to concern themselves with creating additional markup or verifying that the markup was applied correctly. ITS data categories are associated with sets of nodes (for example all p elements in an XML instance)
* Changes can be made in a single location, rather than by searching and modifying local markup throughout a document (or documents, if the [rules](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#rules) element is stored as an external entity)
* ITS data categories can designate attribute values as well as elements.
* It is possible to associate ITS markup with existing markup (for example the term element in DITA)

The commonality in both examples above is the markup translate='no'. This piece of ITS markup can be interpreted as follows:

* it pertains to the [Translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#trans-datacat) data category
* the attribute [translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.local.no-ns.attribute.translate) holds a value of "no"

The ITS [selector](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#att.selector.attribute.selector) attribute allows:

* ITS data category attributes to appear in global rules (even outside of an XML document or schema)
* ITS data categories attributes to pertain to sets of XML nodes (for example all p elements in an XML document)
* ITS markup to pertain to attributes
* ITS markup to [associate with existing markup](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#associating-its-with-existing-markup) (for example the term element in DITA)

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**2.2 Overriding and Inheritance**

The power of the ITS selection mechanisms comes at a price: rules related to [overriding/precedence](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "selection-precedence), and [inheritance](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#datacategories-defaults-etc), have to be established.

The document in [Example 12](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#EX-basic-concepts-3) shows how inheritance and overriding work for the [Translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#trans-datacat) data category. By default elements are translatable. Here, the [translateRule](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#translateRule) element declared in the header overrides the default for the head element inside text and for all its children. Because the title element is actually translatable, the global rule needs to be overridden by a local its:translate="yes". Note that the global rule is processed first, regardless of its position inside the document. In the main body of the document, the default applies, and here it is its:translate="no" that is used to set “faux pas” as non-translatable.

Example 12: Overriding and Inheritance

<text

xmlns:its="http://www.w3.org/2005/11/its" >

 <head>

  <revision>Sep-10-2006 v5</revision>

  <author>Ealasaidh McIan</author>

  <contact>ealasaidh@hogw.ac.uk</contact>

  <title

    its:translate="yes">The Origins of Modern Novel</title>

  <its:rules version="2.0">

   <its:translateRule translate="no" selector="/text/head"/>

  </its:rules>

 </head>

 <body>

  <div xml:id="intro">

   <head>Introduction</head>

   <p>It would certainly be quite a <span

      its:translate="no">faux pas</span>

    to start a dissertation on the origin of modern novel without

    mentioning the <tl>Epic of Gilgamesh</tl>...</p>

  </div>

 </body>

</text>

[Source file: [examples/xml/EX-basic-concepts-3.xml](http://www.w3.org/International/multilingualweb/lt/drafts/its20/examples/xml/EX-basic-concepts-3.xml)]

[**[Go to the table of contents.](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#contents)**2.3 Adding Information or Pointing to Existing Information**

For some data categories, special attributes add or point to information about the selected nodes. For example, the [Localization Note](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html" \l "locNote-datacat) data category can add information to selected nodes (using a [locNote](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#locNote) element), or point to existing information elsewhere in the document (using a [locNotePointer](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#locNoteRule.attributes) attribute).

The functionality of adding information to the selected nodes is available for each data category except [Language Information](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#language-information). Pointing to existing information is not possible for data categories that express *a closed set of values*; that is: [Translate](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#trans-datacat), [Directionality](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#directionality), [Locale Filter](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#LocaleFilter) and [Elements Within Text](http://www.w3.org/International/multilingualweb/lt/drafts/its20/its20.html#elements-within-text).

The functionalities of adding information and pointing to existing information are *mutually exclusive*. That is to say, attributes for pointing and adding must not appear at the same rule element.