POSTSECONDARY
Electronic Standards Council

Recommendation for the adoption of the xml:lang attribute for PESC Schemas

Prepared by the PESC Technical Advisory Board

Revision 1.0.0
7/25/11

Introduction

PESC standards are beginning to find acceptance in countries outside the United States, and U.S. institutions are increasingly recruiting students and faculty internationally. As a result, some of the information provided on transcripts, applications, and other interchange documents may have language content other than U.S. English.

The Technical Advisory Board (TAB) recommends that PESC provide a standard mechanism for identifying the language content of XML elements. By including language information in the XML instance document, processing application can take the appropriate action to deal with language content. For example, the application could route the document or its contents to a human reader to perform a translation, or it could translate Unicode characters that do not map into the database character set into the appropriate representation.

In addition to recommending a mechanism for enabling language identification, this document provides guidance on how to create or modify PESC XML schemas so that data elements may be language enabled.

Recommendation

The TAB recommends that PESC standardize on the xml:lang attribute to define the language content of elements. This attribute is defined as part of XML 1.0 Standard from the W3C.

We recommend that workgroups and XML users groups (e.g., ERUG) identify those elements that may have language content other than U.S. English and add the xml:lang attribute to the element or type definition in the XML schema.

The TAB also recommends that the scope rules for the language attribute follow the W3C recommendation: For elements with simple content, the value of the xml:lang attribute is interpreted to mean that the textual content of the element is in the indicated language. For complex elements with sub-elements, the attribute value applies to all the sub-elements subsumed by that complex element unless overridden by a sub-element xml:lang value.

The W3C has provides further guidance on the use of language tags in XML and HTML in this document at the following URL:

http://www.w3.org/International/articles/language-tags/

This document references all the relevant IETF RFCs as well as the official language and region lists used to make up the language tag. We include this document as part of this recommendation by reference.

Schema Development Guidance

Note that the prefix xml and the xml namespace uri are predefined in XML. XML editors such as XMLSpy will recognize the namespace and make the xml:lang attribute available for reference in the XML schema. To include the XML attribute in an XML Schema requires three steps:

1) Include the xml namespace in the schema element:

```
<xs:schema xmlns:xs=http://www.w3.org/2001/XMLSchema
xmlns:xml="http://www.w3.org/XML/1998/namespace">
```

2) Import the xml: attributes from the xml namespace

```
<xs:import namespace="http://www.w3.org/XML/1998/namespace"/>
```

3) Include the attribute in elements or types that need to be language enabled:

```
<xs:attribute ref="xml:lang" use="optional"/>
```

Below is an example of a simple schema that includes an xml:lang attribute for a complex element and for two simple elements:

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XMLSpy v2009 sp1 (http://www.altova.com) by Michael Morris (Act Inc.) -->
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:xml="http://www.w3.org/XML/1998/namespace" elementFormDefault="qualified"
attributeFormDefault="unqualified">
    <xs:import namespace="http://www.w3.org/XML/1998/namespace"/>
    <xs:element name="TopLevel">
        <xs:complexType>
            <xs:sequence>
                <xs:element name="A">
                    <xs:complexType>
                        <xs:simpleContent>
                            <xs:extension base="xs:string">
                                <xs:attribute ref="xml:lang" use="optional"/>
                            </xs:extension>
                        </xs:simpleContent>
                    </xs:complexType>
                </xs:element>
                <xs:element name="B">
                    <xs:complexType>
                        <xs:simpleContent>
                            <xs:extension base="xs:string">
                                <xs:attribute ref="xml:lang" use="optional"/>
                            </xs:extension>
                        </xs:simpleContent>
                    </xs:complexType>
                </xs:element>
            </xs:sequence>
            <xs:attribute ref="xml:lang" use="optional"/>
        </xs:complexType>
    </xs:element>
</xs:schema>
```

The instance document below is an instantiation of the above schema:

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!--Sample XML file generated by XMLSpy v2009 sp1 (http://www.altova.com)-->
<TopLevel xml:lang="" xsi:noNamespaceSchemaLocation="language%20schema.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
    <A xml:lang="fr">dècembre</A>
    <B xml:lang="it"> Dicembre</B>
</TopLevel>
```

The TopLevel xml:lang empty value is interpreted as “don’t care” while the A and B elements express that the text is in French and Italian respectively.

For those developers who are modifying existing schemas, there is one other issue that arises when adding the language attribute. To add an attribute to a simple type, the simple type must be redefined as a complex type. So even if the name stays the same, a validation error will occur for every element that uses this type. This error can be removed by redefining the element with the new complex type.