BOWNE GLOBAL SOLUTIONS

Introduction to XLIFF

Table of Contents

1.0   Introduction

I am writing this document on a computer built in Ireland by an American company using parts from China and India as well as Europe and America. The same operating system and word processor I am using are to be found in computers throughout the world, in the language spoken by the person using them. We live in a global marketplace where businesses have to work across languages and cultures.

The translation of electronic document, software and Web applications is complicated by, among other things, the range and complexity of the file formats used. XLIFF is designed to make localization easier by providing a standard format for exchanging translatable data.

XLIFF stands for XML Localization Interchange File Format. It was initially developed by a group of companies which included Oracle, Sun, Novell, IBM and Bowne Global Solutions. It has now become an industry standard for localization and is being developed through the OASIS organization.

OASIS is the main body for developing XML-based standards for Industry application. The XLIFF OASIS technical committee, which is overseeing development of XLIFF, has some of the most significant players in the area of software and translation.

This document is aimed at those who work for companies who are using or planning to use XLIFF. The document is structured so that you will gain a clear understanding of the benefits of using XLIFF in the "Why XLIFF" section. "What is XLIFF" will introduce the reader to how this file format is structured and shows what an XLIFF would look like and how it would be used. The case study in the "XLIFF in Action" section show how you would go about starting to use XLIFF.

2.0   Why XLIFF?

2.1 The Need for XLIFF

Work on XLIFF started in 2001 when a number of people from different companies started working together and looking at the issue of file formats and localization. The problem was that, for each software provider, there is a significant number of file formats. The original software provider, the localization provider and the tool vendors had to invest considerable efforts to deal with how to localize each file type. The idea behind XLIFF was to have a standard file format for exchanging localization data.

In 2002, the XLIFF technical committee (TC) was formed at OASIS (Organization for the Advancement of Structured Information Standards) and took over the work of the original group. The key aspects of its design objectives are:

- The file format serves as a container for externalized data to be interchanged between software publishers, localization tools and software service providers in order to facilitate all the phases of the localization process. To achieve this objective, the standard must be tool-neutral.

- The file format must be extensible in order to support new and proprietary data formats. However, it must be structured and well-defined so that tools that support the format would be reliable and consistent.

- End-to-end localization lifecycle support - information relevant to all project phases could be stored within the same file.

The challenges for which the XLIFF TC is aiming to provide solutions include:

- **Many different formats**: Rather than have multiple file formats per project, XLIFF is providing a standard format for exchanging localization data.

- **Version management**: XLIFF offers version management at a segment level.

- **Workflow metadata**: Within the XLIFF file, it is possible to store workflow metadata.

- **Lack of standards**: There has been significant work within the localization industry on a number of standards before XLIFF. However, XLIFF is becoming an industry standard for localization rather than a localization industry standard. There are considerable benefits in a consortium such as OASIS designing standards rather than individual companies re-inventing the wheel.

2.2 Benefits

The different partners involved in a localization project can all benefit from using XLIFF. XLIFF is an XML application and offers these partners the benefits of this technology. These include using XML itself, the ability to transform the file using XSLT and other style sheet languages, and the ability to integrate with Web services.

Using a standardized single format, as opposed to multiple proprietary and other formats, offers many advantages to customers of localization. Quality and training issues are greatly reduced. There is no longer a need to design the localization process for each file format. In the past, many customers of localization developed their own resource editors. With XLIFF, greater reliance can be placed on the independent localization tool and vendors who will want to fully support a more standardized format.

The relationship between customer and localization vendors should be based on the provision of a high-quality, innovative and effective service. With proprietary formats, the choice to find this might be limited because of expertise a particular vendor has in those formats. XLIFF enables the customer to choose based on how the vendor adds value to their products.

In the long term, XLIFF will help all players within the localization industry remain competitive. As the industry uses this standardized format more, resources used to deal with multiple file formats can be focused on adding greater value to the customer's products.

2.3   Where You Would Use XLIFF

XLIFF is designed to be a solution when difficulties arise localizing the file format or exchanging localizable data. The first thing that should be said is that, if you are happy with the speed, efficiency and quality of your localization process, you should be slow to change it. It is good to look at alternative technologies and solutions but there is little point if you are localizing in a file format which has excellent localization tools and processes.

The area where you are unlikely to use XLIFF is where you are localizing common document formats. There are thousands of other formats and sources of data for localization, and it is for these that XLIFF provides the solution. Some of the technologies for which XLIFF is particularly suited include:

* Software in general

* Java applications

* Web applications

* Databases

* HTML documents

* Proprietary formats

In the section "XLIFF in Action," I will return to this subject. I will take you through a number of case studies before looking at how you would go about this with your own file format.

3.0   What Is XLIFF?

XLIFF is an XML application for exchanging localization data. This standard has been written to take full advantage of XML and use this in a file format designed for the localization process. The architecture of XLIFF is designed so that it can be used in the Extract - Localize - Merge process illustrated below.
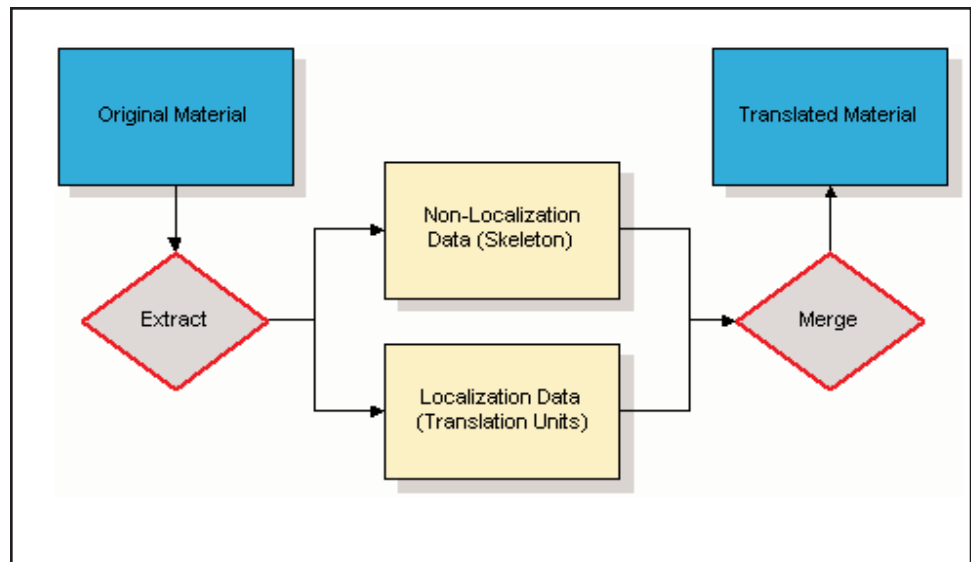
Extract-Localize-Merge

The original source files to be localized are filtered with the localization data extracted into the XLIFF file. With complex file formats such as software, the non-localization data can be stored as a skeleton file. This can be done within the XLIFF file or externally with a reference to this within the XLIFF file. On completion of the translation process, the XLIFF is merged with the non-localization data to create a localized version of the original file.

3.1 Structure of XLIFF

The following is an example of a simple XLIFF file.

```xml
<?xml version="1.0" encoding="utf-8"?>
<xliff version="1.1">
<file source-language='en' target-language='fr' datatype="database" original="LR GWS">
<header>
    <count-group>
     <count count-type="word">11</count>
    </count-group>
</header>
<body>
<trans-unit>
     <source xml:lang="EN-US">We are testing XLIFF.</source>
     <target xml:lang="FR">Nous testons XLIFF.</target>
</trans-unit>
<trans-unit>
     <source xml:lang="EN-US">The translation portal.</source>
     <target xml:lang="FR">Le portal de traduction.</target>
</trans-unit>
</body>
</file>
</xliff>
```

This document does not intend to describe in detail all the components which make up an XLIFF file. For further information on this, refer to the XLIFF specification from OAISIS. The current specification is available at http://www.oasis-open.org/commit-tees/xliff/documents/cs-xliff-core-1.1-20031031.htm .

XLIFF files are bilingual. There is only one source language and one target language. This makes the overall model much simpler when you consider all the additional features the format provides, such as pre-translation, version tracking, etc. It also avoids the problem which a multilingual file would have where different versions are localized by different translators at the same time.

3.1.1 file element

Each file element corresponds to a real or virtual file from the source material. By virtual file, I mean that it could correspond to part of a database. An XLIFF file must contain one but can contain any number of file elements. It would be possible to have a number of resource files, used in a software application, that each corresponded to a file element in the one XLIFF file. You should take into consideration the effect that having many file elements will have on the size of the XLIFF file and managing the project.

The file element has a number of attributes which contain data describing the actual file to which they relate.

Each file element contains both a header and body child elements. The header element contains meta data about the file being localized. The body element contains the localization data.

3.1.2 trans-unit

The trans-unit is where the data to be localized is contained. Each trans-unit has a source element which contains the original data for localization. After translation, the target element contains the data which has been localized.

```
<trans-unit>
     <source xml:lang="EN-US">We are testing XLIFF.</source>
     <target xml:lang="FR">Nous testons XLIFF.</target>
</trans-unit>
```

The attributes of the trans-unit element can be used to store data such as the type of resource being translated. The box below shows a cancel button. Using this information, a localization resource editor can render the button or windows resource which can then be localized in a WYSIWYG editor.

```
<trans-unit id='1' resname='IDCANCEL' restype='button' coord='8;8;50;14'>
    <source xml:lang='en'>Cancel</source>
</trans-unit>
```

The trans-unit elements can be grouped together using the group element. An example where you might want to do this is when you are localizing a dialog box and you are grouping all the different components of the dialog box.

3.1.3 bin-unit

The bin-unit is used to store non-textual data such as icons, cursors or images. The object can be stored as an external file or stored internally. When stored as an external file, the bin-unit contains the address where the file is found. The binary file can also be stored within the bin-unit as base-64 encoded data.

If there is localizable text associated with the binary object, it can be stored in one or more trans-unit elements contained within the bin-unit.

3.2 Extensibility

One of the major initiatives within XLIFF 1.1 is the use of XML schema namespaces to allow XLIFF to be embedded within another XML document or for XLIFF to be extended with another namespace. A company might have an XML application which meets their needs in all areas except localization. A small section of XLIFF could be embedded within this file to deal with that. It is also possible to embed a section of another XML application at particular extension points within an XML document.

An example of how an XLIFF document could be extended is shown below. The 'mysc' elements are used within the XLIFF document by using this namespace.

```
<xliff version='1.1'
 xmlns='urn:oasis:names:tc:xliff:document:1.1'
 xmlns:mysc='http://www.myschema.com/myschema-v1'>
 <file original='mydoc.doc' source-language='en' traget-language='fr' datatype='plain-text'>
  <group>
   < mysc:Info>
    <mysc:deatails>details go here</sup: deatails >
   </ mysc:Info>
   <trans-unit id='1'>
    <source xml:lang='en'>Text to be translated </source>
    <target xml:lang='fr'>translated text</target>
   </trans-unit>
  </group>
 </file>
</xliff>
```

3.3 Alternate Translations

The alt-trans element is used to store alternative translations. This could be Translation Memory, Machine Translation, a previous translation or a previous translation in a different dialect to the current target language. An example of where you would store a different dialect would be where you have translated an English document to French Canadian and you previously have a French translation. The XLIFF file would have English as the source language and French Canadian as the target language but you might have the French as the alternative translation in the alt-trans element.

The example below illustrates a previous translation of similar text is available and the alt-trans element, with this translation, is placed within the trans-unit.

```
<trans-unit>
      <source xml:lang="EN-US">We are testing XLIFF.</source>
      <target xml:lang="FR">Nous testons XLIFF.</target>
      <alt-trans>
          <source xml:lang="FR">We are testing.</source>
          <target xml:lang="FR">Nous testons.</target>
      <alt-trans>
</trans-unit>
```

3.4 Version Control

Within an XLIFF document, it is possible to keep track of what is happening with the use of the phase element. This can be used for version control where different versions of the translation are stored in the alt-trans element or it can used to provide other control data.

The phase element is used inside the header element and this provides the details of a particular phase. Information such as what was done to a segment with what tool can be stored here.

```
<header>
 <phase-group>
  <phase phase-name='trans' process-name='translation' tool='Trados' date='2005-04-01T20:01:23Z'/>
  <phase phase-name='edit' process-name='Editing' tool='Trados' date='2005-04-01T20:01:23Z'/>
 </phase-group>
</header>
```

A target element can now be associated with a particular phase. In the example below, we have used the phase element to show that the target element was from the phase 'trans'. Details of this phase are found in the header section.

```
<trans-unit>
     <source xml:lang="EN-US">We are testing XLIFF.</source>
     <target xml:lang="FR">Nous testons XLIFF.</target>
     <alt-trans>
          <source xml:lang="FR">We are testing.</source>
          <target xml:lang="FR" phase-name='trans'>Nous testons.</target>
     <alt-trans>
</trans-unit>
```

4.0   XLIFF in Action

4.1   XLIFF in Action with Web Services

This solution was implemented for a customer of Bowne Global Solutions who is a large car manufacturer. They wanted to increase their presence in the global market. They had a significant Web presence but wanted to expand their Web site and the range of languages available on their site. There was a complicated process involved in the localization of their existing Web site. They were concerned that this process was adding extra cost without the corresponding value.

Together with this customer, Bowne Global Solutions designed a solution which made innovative use of XLIFF, Web Services and BGS's online translation service, ElcanoTM. The diagram below shows the system which was built in order to achieve this. The repository on the customer's side was a Content Management System.

Elcano web services overview
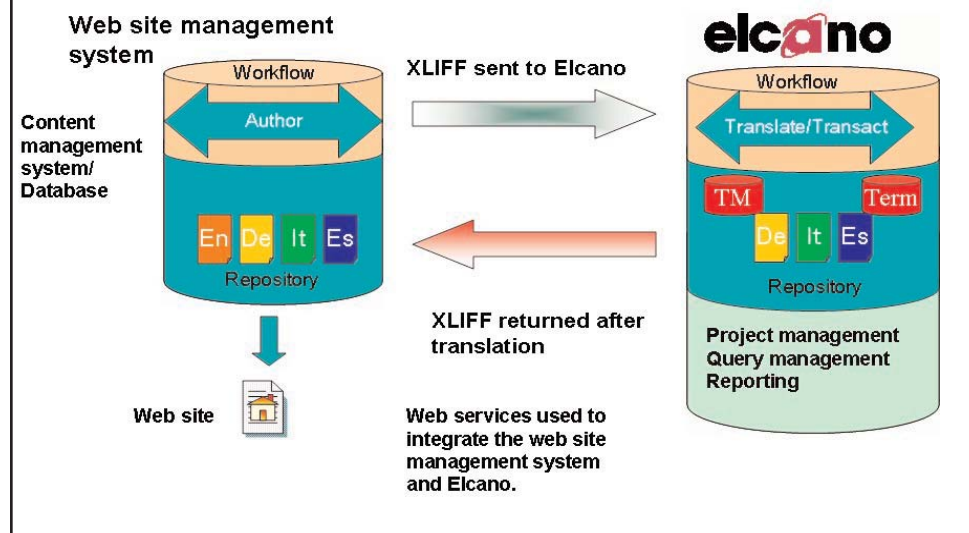
The XLIFF files which were transported to and from the Content Management System were translated using Trados TagEditor. TagEditor is a translation tool for XML, and it uses an ini file to store details of Trados TagEditor should work with a particular file format. An ini file was created to work with XLIFF that was a modification of one that Trados make available on their Web site. This ini file prevents the translator from editing any of the XLIFF mark-up, the source data or anything which they should not edit. This tool also directly works with Trados Translation Memory technology.

What was striking during this project was how stable XLIFF was as a localization format.

At that time, we needed to localize problematic HTML files. Using XLIFF assured the technical quality of the end delivery.
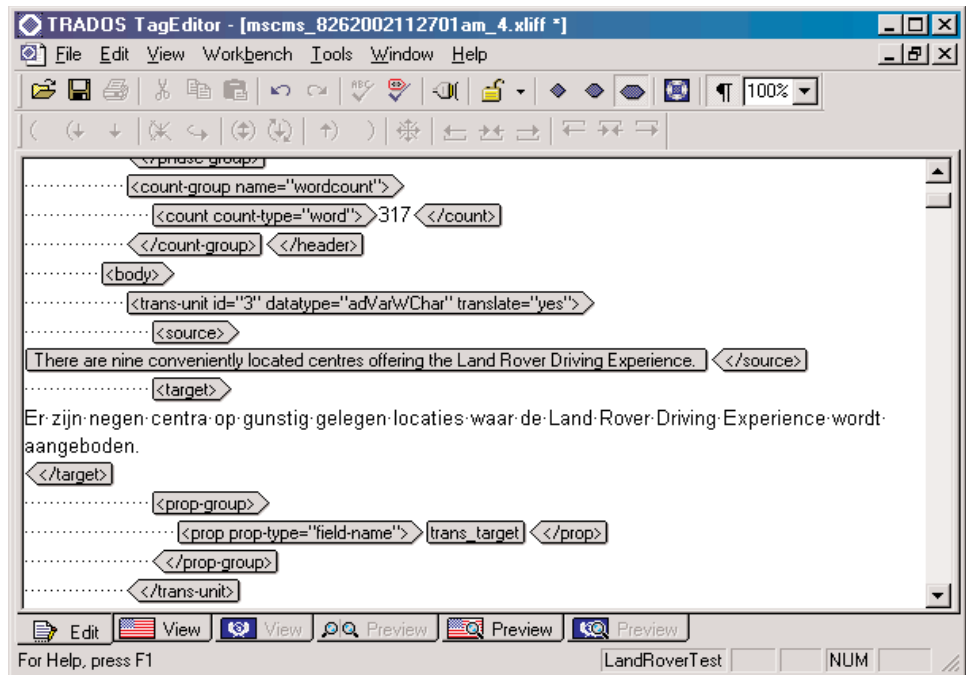
4.2 Software Examples

XLIFF has the capacity to store the data which is used to render a Windows resource. A number of resource editors, such as Catalyst from Alchemy Software or PASSOLO from PASS Engineering, provide WYSIWYG editing. The XLIFF file below shows an example where the data for a Windows menu is included. This enables the localization resource editor to show this Windows resource as it would appear on a screen and for WYSIWYG editing to be used.

The example below shows how the trans-unit elements for a Windows menu are represented.
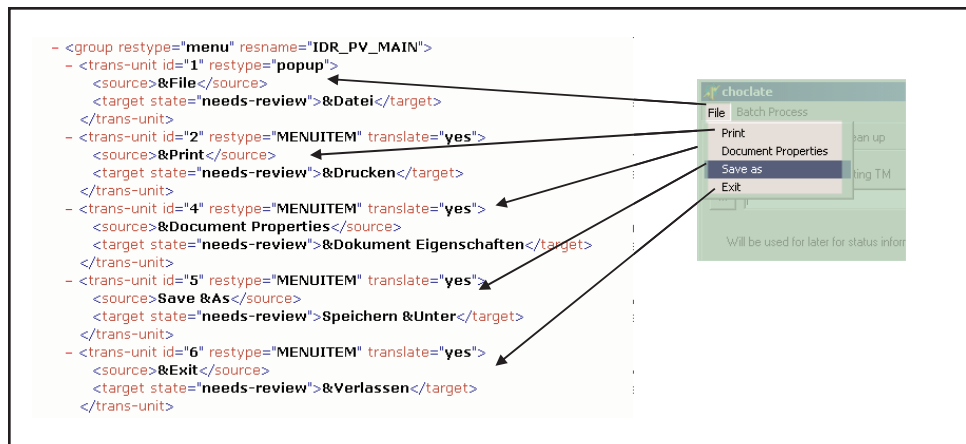
4.3 Database/CMS Examples

XLIFF can be used to extract localization data from a database. This data would be localized within the XLIFF file before being imported back into the database. Before XLIFF, the methods for extracting, translating and importing localization data from a database would have been quite cumbersome. This could have involved creating a proprietary file format or using a copy of the database.

```
<group id="234432" resname="DATABASE">
   <group id="22" resname="TABLE">
   <group id="441" resname=" TABLE_RECORD">
       <trans-unit id='1'>
         <source xml:lang='en'>Source text for title</source>
         <target xml:lang='fr'>translated text for title</target>
       </trans-unit>
       <trans-unit id='1'>
         <source xml:lang='en'>Source text for description</source>
         <target xml:lang='fr'>translated text for description</target>
       </trans-unit>
    </group>
    </group>
</group>
```

4.4 Steps to Using XLIFF

The following is intended as advice for someone whose company is considering using XLIFF within their localization process. It should be considered as a starting point for your decision on whether or not to use XLIFF rather than a comprehensive guide.

Will XLIFF Add Value?

If you are already using a file format which is well supported by your localization process and tools, then I would suggest that you should be careful before you change. I would strongly advise any company considering XLIFF to evaluate it thoroughly. If possible, an analysis should be carried out from both a technology and a business perspective.

XLIFF will add very significant value in most circumstances but especially when you are dealing with a complex format or when the process and tools are not clearly defined.

Establish Your Process

The diagram below was shown previously at the beginning of the section "What is XLIFF."
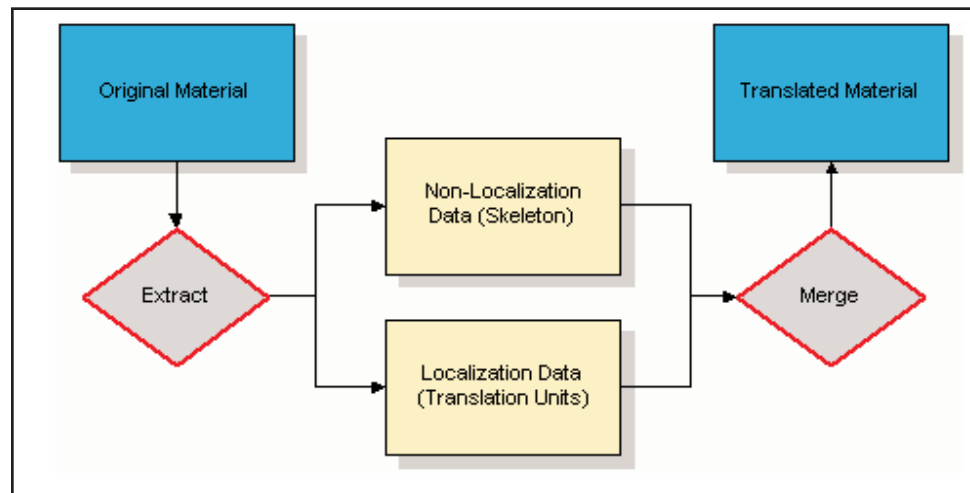
A company establishing their process for using XLIFF will need to decide on a number of factors relating to the Extract-Localize-Merge process.

♦ What will be the mechanism to extract the XLIFF file from the original material. If the original material is XML-based, the company can make use of XSLT to transform the file to XLIFF.

♦ There are choices mentioned above on how to construct an XLIFF document. I would suggest that this should be kept simple with one file element per XLIFF file. If a skeleton file is used, a decision needs to be made on whether this is stored internally or externally.

♦ While the company does not have to prescribe a particular tool, it is recommended that the tool being used can get the full benefits of using XLIFF. For example, a localization resource editor such as Catalyst from Alchemy Software or PASSOLO from PASS Engineering provide WYSIWYG editing of an XLIFF file. There is also support for XLIFF in the tools built by Trados, Heartsome and RC-WinTrans from Schaudin.com.

♦ As with the Extract stage, a filter needs to be used to create the translated material by converting the XLIFF file at the merge stage.

How to Implement

After you have established the advantages of XLIFF and have evaluated and decided on your process you are now ready to start using XLIFF to localize your files.

The first step in implementation is to explain your process to the stakeholders. Those involved in the creation of the original files and those involved at all stages of the localization process should all be explained as to why you are moving to XLIFF and how you intend doing this.

It would be wise to start with a pilot project. This would give you a chance to evaluate the complete process initially on a small scale.

As part of evaluating the process you will have decided on the tools you are using for the extracting and importing to and from XLIFF. You can also decide whether you want to make recommendations on the tools used for translation. It is good practise to produce a project kit to support those involved in the localization. This would include style sheets and supporting documentation. You should include information on XLIFF, why you are using it and any recommendation.

The engineers responsible for extracting the XLIFF would prepare the XLIFF files and these will be sent for localization. The files would then be translated and re-imported to create localized versions of the original files.

Following the success of the pilot project you can start using XLIFF for more of your localization.

5.0   BGS and XLIFF

As mentioned earlier in this white paper, in September 2000 a number of companies involved in the localization industry started working together. Software localization was becoming more complex with many different file formats being localized for any given project. Their efforts to find a solution to these issues led to the development of XLIFF. Bowne Global Solutions (then Berlitz GlobalNET) was one of these companies along with Novell, Oracle, IBM and Sun. We have been active in the technical committee (TC) designing XLIFF since then.

The original XLIFF technical committee decided that the best way to progress further was to do this through OASIS. In December 2001 there was a call for participation in the XLIFF TC. Peter Reynolds, who works as a Software Development Manager with BGS, has been secretary of the XLIFF TC since it joined OASIS.

From an early stage BGS was very enthusiastic about what XLIFF had to offer our customers. We had agreed to sponsor the Special Olympic World Games in Ireland in 2003. This was the first time the games were held outside of North America and thanks to BGS's sponsorship was the first time that the Web site for games was translated into dif-

ferent languages. BGS also developed an innovative way to send short stories to BGS for translation and back to the Web site. When a new story was submitted, the editor could choose which languages it was also to be translated in. An XLIFF file was then created for each of these languages and this would be sent to BGS's online translation service, Elcano, which managed the translation.

The model of using XLIFF in conjunction with Web Services was used successfully by BGS. The XLIFF in action with Web Services section above (4.1) provides more information on this.

BGS's role as the largest localization vendor has meant that many of our customers have been early adopters of XLIFF. We have worked with them in localizing their products, ensuring that they are internationalized, and helping streamline their process and move towards using XLIFF.

6.0 Reference

6.1 OASIS

OASIS (Organization for the Advancement of Structured Information Standards) is a not-for-profit, international consortium that drives the development, convergence and adoption of e-business standards. The consortium produces more Web Services standards than any other organization along with standards for security, e-business, and standardization efforts in the public sector and for application-specific markets. Founded in 1993, OASIS has more than 4,000 participants representing over 600 organizations and individual members in 100 countries. http://www.oasis-open.org/

6.2 XLIFF Technical Committee

The XLIFF TC at OASIS comprises representatives from a wide range of companies including software development companies, localization vendors and localization tool vendors. Anyone who is a member of OASIS can participate in these meetings. If you are interested in joining the TC, please contact Peter Reynolds, the TC secretary, at peter.reynolds@bowneglobal.ie or Tony Jewtushenko, chair of the TC, at tony.jewtushenko@productinnovator.com.

6.3 Web References

IBM - Globalize Your on Demand Business - A white paper on using XLIFF. http://www.306.ibm.com/software/globalization/topics/xliff/index.jsp

XML in Localization - Use XLIFF to translate documents. http://www.128.ibm.com/developerworks/xml/library/x-localis2/

Using Translation Technology at Sun Microsystems - A white paper which describes Sun's experience of using XLIFF. http://developers.sun.com/dev/gadc/technicalpublications/whitepapers/translation_technology_sun.html

Introduction to XLIFF - Tony Jewtushenko, chair of XLIFF TC. http://www.localisation.ie/publications/presentations/2004/SummerSchool2004/Presentations&files/Tony%20Jewtushenko/LRCSS04_XLIFF_overview.ppt

Mulilingual Magazine - Articles on XLIFF - On the menu go to Magazine then featured articles. There is an article by Yves Savourel - An Introduction to Using XLIFF - in Volume 14 Issue 2 and an article by ULTAN O BROIN - Internationalization Features in XML and XLIFF - in Volume 13 Issue 2.. http://www.multilingual.com/

BGS would welcome the opportunity to explore the value our enterprise-class solutions can bring to your organization and would invite you to contact us for an initial discussion about your global objectives and how we might be able to help you achieve them.

To learn more, contact us at info@bowneglobal.com or find us on the web at www.bowneglobal.com.

About BGS

Bowne Global Solutions (BGS) is the leading provider of translation, localization, technical writing and interpretation services that enable businesses to deliver locally relevant and culturally connected products, services and communications anywhere in the world. Companies throughout the world use our solutions to help grow their businesses in the Americas, Europe, Asia and Latin America.

Our scalable end-to-end solutions can help accelerate a company's time to market while improving the quality and consistency of the company's products and services. Our wide range of clients include leading businesses in information technology, automotive, e-learning, life sciences, entertainment, telecommunications, aerospace and power and utilities industries.

Bowne Global Solutions was formed in 1997 through the acquisition of IDOC, GECAP, Pacifitech, ME&TA and I&G Com. We acquired Mendez, a leading provider of globalization solutions in Europe, in 2001, followed by Berlitz GlobalNET, another leader in the industry, in 2002. Our expert professionals are located in more than 40 production and project management sites in 24 countries and have access to a worldwide network of 10,000 qualified linguistic resources covering more than 60 languages and dialects.

BGS, which has its global headquarters in New York, is a business unit of Bowne & Co., Inc. (NYSE: BNE), a global leader in providing high-value solutions that empower our clients' communications. In addition to BGS, Bowne & Co.'s diversified business units include Bowne Financial Print, the world's largest financial printer; Bowne Enterprise Solutions, a leading provider of digital printing and electronic delivery of communications; and Bowne Business Solutions, the leader in delivering a full array of business process outsourcing services to the legal and financial services industries.

To learn more, contact us at info@bowneglobal.com or find us on the web at www.bowneglobal.com.