

An OASIS DITA Adoption Technical Committee Publication

DITA 1.2 Feature Article: Acronym Best Practices

DITA Translation Subcommittee
On behalf of the OASIS DITA Adoption Technical Committee

Date: 24 May 2010



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. Members themselves set the OASIS technical agenda, using a lightweight, open process expressly designed to promote industry consensus and unite disparate efforts. The consortium produces open standards for Web services, security, e-business, and standardization efforts in the public sector and for application-specific markets. OASIS was founded in 1993. More information can be found on the OASIS website at <http://www.oasis-open.org>.

The OASIS DITA Adoption Technical Committee members collaborate to provide expertise and resources to educate the marketplace on the value of the DITA OASIS standard. By raising awareness of the benefits offered by DITA, the DITA Adoption Technical Committee expects the demand for, and availability of, DITA conforming products and services to increase, resulting in a greater choice of tools and platforms and an expanded DITA community of users, suppliers, and consultants.

DISCLAIMER: All examples presented in this article were produced using one or more tools chosen at the author's discretion and in no way reflect endorsement of the tools by the OASIS DITA Adoption Technical Committee.

This white paper was produced and approved by the OASIS DITA Adoption Technical Committee as a Committee Draft. It has not been reviewed and/or approved by the OASIS membership at-large.

Copyright © 2010 OASIS. All rights reserved.

All capitalized terms in the following text have the meanings assigned to them in the OASIS Intellectual Property Rights Policy (the "OASIS IPR Policy"). The full Policy may be found at the OASIS website. This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published, and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this section are included on all such copies and derivative works. However, this document itself may not be modified in any way, including by removing the copyright notice or references to OASIS, except as needed for the purpose of developing any document or deliverable produced by an OASIS Technical Committee (in which case the rules applicable to copyrights, as set forth in the OASIS IPR Policy, must be followed) or as required to translate it into languages other than English. The limited permissions granted above are perpetual and will not be revoked by OASIS or its successors or assigns. This document and the information contained herein is provided on an "AS IS" basis and OASIS DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY OWNERSHIP RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Document History

Revision	Date	Author	Summary

Table of Contents

Statement of the Acronym Problem	7
Recommended Best Practices	8
Special conditions related to the translation of acronyms	9
Different forms in the source and target languages	9
Potential for grammar errors	9
Problems with inflected languages	10
Processing instructions	10
Instruction to the translators	11
Translating the glossary entries	11



Statement of the Acronym Problem

Abbreviated forms such as acronyms are used frequently in technical documentation. Abbreviated forms need to be expanded to their full form the first time that they appear in a document to ensure that the reader understands what the abbreviated form refers to. In electronic published documents such as an online help system, the expansion of abbreviated forms can also be made available in the form of a hyperlink or 'tool tip' mechanism. In addition, it should be possible to automatically insert the expansion of abbreviated forms from the source file into glossary entries for the publication. This best practice describes how to encapsulate abbreviations and their full forms in DITA documents to realize these objectives.

Abbreviated forms and their translations require special handling. Some abbreviated forms are never translated, especially those that are intended for a knowledgeable, technical audience, and those that refer to standardized international concepts, such as "XML". Some abbreviated forms represent a brand name for which the original expanded form is no longer used or is used less frequently than the abbreviated form. Some abbreviated forms such as xml, jpg, html, and so on are typically used in their original lower case form, while normally acronyms are used in upper case.

Abbreviated forms may or may not have a corresponding abbreviated form in a given target language. For example, United Nations (UN) and Weapons of Mass Destruction (WMD) have equivalents in other languages, such as "ONU" and "ADM" for French.

Some English abbreviated forms are retained in the target language for universal recognition purposes and to facilitate search, but the corresponding full form is also provided in a translated version so that the reader understands what the abbreviation means. For instance, "OASIS" may be used unchanged in a translated document, but its translated full form may be included as well (such as "Organisation pour l'avancement des normes sur l'information structure").

Some English abbreviated forms are retained in the target language for universal recognition purposes and to facilitate search, but the corresponding full form is also provided in a translated version so that the reader understands what the abbreviation means. For instance, "OASIS" may be used unchanged in a translated document, but its translated full form may be included as well (such as "Organisation pour l'avancement des normes sur l'information structure").

For example, the first occurrence of an abbreviated form in English might consist of the abbreviated form followed by its expanded form in parentheses. By contrast, the translated version might consist of the expanded form followed by the abbreviated form in parentheses. The translated version might also include both the English text and the translation.

For example, in Polish, the first reference to JSP may appear as follows:

"JSP (ang. Java Server Pages)"

Also in Polish, the OASIS acronym may appear as follows:

"OASIS (ang. Organization for the Advancement of Structured Information Systems - organizacja dla propagowania strukturalnych systemów infomracyjnych)"

In the first example, the translator assumes that the reader will not require a translation of the English expanded form. In the second example, the translator assumes that the reader may not understand the English expanded form and so adds the translation.

To address these requirements for translated text, the DITA 1.2 glossary and acronym specialization assists in the resolution and handling of abbreviated-form text such as acronyms, general abbreviations, and short forms in source and target text within DITA documents.

Recommended Best Practices

To properly represent abbreviations in a DITA document, you use the glossary specialization, creating one or more collection topics to hold abbreviations and their expansions. You may declare an acronym with a glossentry topic similar to the following example:

```
<glossentry id="abs">
  <glossterm>Anti-lock Braking System</glossterm>
  <glossBody>
    <glossSurfaceForm>Anti-lock Braking System (ABS)</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>ABS</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

The `<glossterm>` declares the expanded form of the acronym. The `<glossAcronym>` declares the abbreviated form that you will use in the text. The `<glossSurfaceForm>` shows how the term must appear in the first instance of a printed document or as a tool tip or other representation in an online document. The use of the `<glossSurfaceForm>` allows translators to offer variations on the presentation of an acronym expansion as required in the target language. For example, some languages require that the abbreviated form in parentheses be placed before rather than after the expanded form. Some languages do not use parentheses to contain abbreviated forms. For example, Spanish sometimes placed the abbreviated form in commas rather than parentheses.

The `<glossSurfaceForm>` has been added to account for target languages that render the first occurrence differently than the rendering in the source language.

You then declare a key for the acronym using the standard DITA 1.2 keyref mechanism:

```
<map>
...
<topicref href="maintcar.dita"/>
...
<glossref keys="abs" href="antiLockBrake.dita"/>
... key declarations for other referenced acronyms ...
</map>
```

You can then refer to the acronym using the standard DITA 1.2 keyref mechanism:

```
<task id="maintcar">
...
  <info>The <abbreviated-form keyref="abs"/> will prevent the car from skidding ...</info>
...
</task>
```

For instance, if the topic with the keyref to the “abs” key provided the first occurrence of the ABS term in a printed document, the sentence could be rendered as follows:

“The Anti-lock Braking System (ABS) will prevent the car from skidding in adverse weather conditions.”

If the ABS term had occurred previously within the document, the same sentence could instead be rendered as follows:

“The ABS will prevent the car from skidding in adverse weather conditions.”

Note that the keyref value does not need to match the acronym. In fact, using a value for the keyref that is more likely to be unique will reduce conflicts in situations where the one acronym corresponds to multiple full forms. For example, one could use “cars.abs” as the key for Anti-lock Braking System and “ship.abs” to refer to the American Bureau of Shipping.

Special conditions related to the translation of acronyms

The following cases must be considered for documents that require translation:

Different forms in the source and target languages

A term that has an abbreviation in the source language may not have an abbreviation in the target language and vice-versa. The preferred term may be the abbreviation in the source language, or it may be the full form in the target language and vice-versa.

Note that Computer Assisted Translation (CAT) tools do not allow the translator to change the XML markup. For that reason, you must provide all the glossentry elements in the source language so that they may be omitted or used in a target language as necessary while preserving the markup structure.

The following example illustrates this approach for an English glossary entry topic:

```
<glossentry id="wmd" xml:lang="en">
  <glossterm>Weapons of Mass Destruction</glossterm>
  <glossBody>
    <glossSurfaceForm>Weapons of Mass Destruction (WMD)</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>WMD</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

In Spanish, there is no abbreviation in use for “Weapons of Mass Destruction.” As a result, the `<glossSurfaceForm>` may be left empty.

```
<glossentry id="wmd" xml:lang="es">
  <glossterm>armas de destrucción masiva</glossterm>
  <glossBody>
    <glossSurfaceForm></glossSurfaceForm>
    <glossAlt>
      <glossAcronym></glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

Term resolution processing should always ignore empty elements. If the `<glossAcronym>` and `<glossSurfaceForm>` elements are empty, an `<abbreviated-form>` reference should resolve to the `<glossterm>` text. Thus, if allowed by the CAT tool, the translator can leave the `<glossAcronym>` and `<glossSurfaceForm>` elements empty. The automatic processing of the empty elements should produce the same effect as if the translator had copied the `<glossterm>` text into the empty elements.

However, some CAT tools may not permit the translator to leave an element empty, if it is not also empty in the source language, and will generate an error message that the translation is incomplete. In that case, the translator must duplicate the `<glossterm>` into the `<glossAcronym>` and `<glossSurfaceForm>` elements.

```
<glossentry id="wmd" xml:lang="es">
  <glossterm>armas de destrucción masiva</glossterm>
  <glossBody>
    <glossSurfaceForm>armas de destrucción masiva</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>armas de destrucción masiva</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

Potential for grammar errors

In some languages, such as Spanish, the expansions of abbreviated forms should be written in lower case. If such a lower-case term is automatically inserted, through the keyref mechanism, at the beginning of a sentence, this would incorrectly result in a sentence starting with a lower case character. Depending upon the translation environment and the specific target language requirements, the translator may have to remove the keyref with the expansion and correctly capitalize the first character of the expansion in the sentence.



For example, the acronym for AIDS should be represented as follows in Spanish:

```
<glossentry id="aids" xml:lang="es">
  <glossterm>síndrome de inmunodeficiencia adquirida</glossterm>
  <glossBody>
    <glossSurfaceForm>síndrome de inmunodeficiencia adquirida (SIDA)</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>SIDA</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

Normally the `<glossSurfaceForm>` text from the above example could not be inserted by using a keyref at the beginning of a sentence, because it begins with a lower case letter.

Errors can also occur with preceding articles, such as “a” and “an” in English. The English writer may correct the error before the file is sent for translation. Depending upon the translation environment and the specific target language requirements, the translator may have to remove the keyref and then correctly translate the sentence with either the abbreviated or the expanded form in place.

Problems with inflected languages

Abbreviated forms can cause problems for inflected languages because their expanded form needs to be presented in the nominative case, without any inflection. This gender-neutral form can be achieved with a surface form that provides the full form in parentheses immediately following the acronym.

For example, the Polish acronym for the European Union is:

```
<glossentry id="eu" xml:lang="pl">
  <glossterm>Unia Europejska</glossterm>
  <glossBody>
    <glossSurfaceForm>UE (Unia Europejska)</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>UE</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

Using the above construct enables automated handling of the abbreviated form in Polish without causing any problems with grammatical inflection in running text.

For example, if we were stating that something occurred within the EU, in Polish the locative case would be required: “Unii Europejskiej”, instead of the form in the glossentry: “Unia Europejska”.

But if we were using the abbreviated form instead, it would be invariable in running text, because abbreviated forms are not inflected.

For example the phrase “In the European Union (EU), there are many institutions...” would be translated as follows in Polish:

“W Unii Europejskiej (UE) jest wiele instytucji...”

Whereas by allowing the translator to control how the text is displayed in the `<glossSurfaceForm>`, we can put the abbreviation first :

“W UE (Unia Europejska) jest wiele instytucji...”

Processing instructions

Processors should resolve the keyref to the `<glossSurfaceForm>` in the first occurrence of the term in a printed document and to the `<glossAcronym>` in subsequent occurrences. Likewise, the processors may resolve the keyref using a tool tip or other form in an online document. For example, for the “Anti-lock Braking System,” processes should resolve the “ABS” reference to “Anti-lock Braking System (ABS)” in the first occurrence in a printed document or as a tool tip or other form in an online document and to “ABS” in subsequent occurrences.

If the `<glossAcronym>` is empty because no acronym exists in the target language, the processor must resolve to the `<glossterm>`.

Instruction to the translators

Translations should follow these special instructions in handling acronyms.

Translating the glossary entries

The following examples show how the glossary entries should be translated in various situations. The examples use one term and the French language for demonstrative purposes and are not meant to represent actual usage in French.

The examples use the following typical glossary entry for an English acronym:

```
<glossentry id="abs">
  <glossterm>Anti-lock Braking System</glossterm>
  <glossBody>
    <glossSurfaceForm>Anti-lock Braking System (ABS)</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>ABS</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

Example 1. The two languages are parallel, that is, there is an acceptable translation of the English full form and of the English abbreviation, and the preferred representation for the first occurrence follows the same order in both languages.

```
<glossentry id="abs">
  <glossterm>système de freinage antiblocage</glossterm>
  <glossBody>
    <glossSurfaceForm>système de freinage antiblocage (SFA)</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>SFA</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

Example 2. The English abbreviation is used in the target language.

```
<glossentry id="abs">
  <glossterm>système de freinage antiblocage</glossterm>
  <glossBody>
    <glossSurfaceForm>système de freinage antiblocage (ABS)</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>ABS</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

Example 3. There is no abbreviation in the target language, and the English abbreviation would not be recognized.

In this case, do not include any abbreviation in `<glossSurfaceForm>` and leave the `<glossAcronym>` element empty.

```
<glossentry id="abs">
  <glossterm>système de freinage antiblocage</glossterm>
  <glossBody>
    <glossSurfaceForm>système de freinage antiblocage</glossSurfaceForm>
    <glossAlt>
      <glossAcronym></glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```

If your CAT tool does not support leaving the `<glossAcronym>` element empty, put the full form in it, as follows:

```
<glossAcronym>système de freinage antiblocage</glossAcronym>
```

Example 4. It is preferable to put the abbreviated form first in the target language, because it is more commonly recognized or to avoid required adjustments for inline resolution.

```
<glossentry id="abs">
  <glossterm>système de freinage antiblocage</glossterm>
```



```
<glossBody>
  <glossSurfaceForm>(SFA) système de freinage antiblocage</glossSurfaceForm>
  <glossAlt>
    <glossAcronym>SFA</glossAcronym>
  </glossAlt>
</glossBody>
</glossentry>
```

Example 5. The English abbreviation is used in the target language, as well as its full form. A translation of the full form is needed for clarification purposes on the first occurrence.

```
<glossentry id="abs">
  <glossterm>Anti-lock Braking System</glossterm>
  <glossBody>
    <glossSurfaceForm> Anti-lock Braking System (ABS - système de freinage antiblocage)</glossSurfaceForm>
    <glossAlt>
      <glossAcronym>ABS</glossAcronym>
    </glossAlt>
  </glossBody>
</glossentry>
```