

An OASIS DITA Adoption TC White Paper

DITA 1.2 Feature Description: Improved glossary and terminology handling

Tony Self
On behalf of the OASIS DITA Adoption Technical Committee

Date: 16 December 2009



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 © 2009 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.

Table of Contents

Improved glossary and terminology handling.....	5
Purpose of a glossary.....	5
Purpose of the change to glossary-related elements in DITA 1.2.....	5
Benefit to users.....	5
The glossentry information type.....	5
How a glossary entry might be rendered.....	7
Incorporating a glossary in a ditamap.....	8
Creating a glossary section.....	9
Linking terms in content topics.....	9
About keys and keyref attributes, and indirect linking.....	9
Summary.....	10

Improved glossary and terminology handling

DITA 1.2 introduces significant improvements to the way in which terminology is handled. A new glossary entry element provides a flexible semantic structure to store terminology information, and the new indirect linking functionality allows terms occurring in the content to be linked through a key reference or transcluded using a new abbreviated-form element. Keys used for key referencing are defined in a new glossref specialisation of topicref in the map.

Purpose of a glossary

The purpose of a glossary is to allow readers who encounter terms that they do not understand to look up a definition for the term. Terms may be single words or longer phrases, they may be acronyms or abbreviations, or they may be domain-specific jargon. In a printed publication, a glossary (usually referred to as a *Glossary of Terms*), appears at the front or back of the book content. In an online document, the glossary is often accessed from a link on the document toolbar and typically displays in a separate window or frame from the document content proper. Many online documents make the glossary easier to access by hyperlinking terms in the content to the associated definitions in the glossary.

Purpose of the change to glossary-related elements in DITA 1.2

The new elements introduced in DITA 1.2 will make it easier for publishing tools to automatically hyperlink terms appearing in the text with their respective definitions in the glossary (in online document outputs). The new elements will also make it easier for users to semantically identify terms and their definitions through a new glossentry information type.

The glossentry is a specialisation of the concept information type.

A new glossref element has also been introduced to provide for the possibility of linking terms in the document content with the glossary definitions.

There are three distinct processes in setting up a glossary in DITA.

- Creating individual glossary topics (using the glossentry information type)
- Incorporating the glossary topics into a ditamap
- Linking terms in content topics to their glossary definitions.

Benefit to users

The new glossary-related elements will open up opportunities, through improvements in publishing tools, for the following features in output documents:

- Automated pop-up or expansion linking of terms in the content to definitions in the glossary.
- Generation of a rich and comprehensive Glossary of Terms for a printed document.
- Generation of a dynamic Glossary of Terms (perhaps sortable and searchable) for an online document.

Another benefit will be the ability to use a term in an abbreviated form (for example, an acronym) or in a longer form, yet still identify the relationship between the term and its definition.

The glossentry information type

The glossentry information type contains an individual glossary entry which is made up of the following elements.

Element	Element Name	Provides	Example
Glossary term	glossterm	The preferred form of the term.	<code><glossterm> USB flash drive </glossterm></code>
Definition	glossdef	The definition or explanation of one sense of the term.	<code><glossdef> A small portable drive. </glossdef></code>
Body of the glossary	glossBody	Container for more details about a glossary term (such as part of speech or additional forms of the term)	<code><glossBody><... /></glossBody></code>

In turn, the optional `glossBody` element contains the following elements, used to store more detailed information about the glossary entry.

Element	Element Name	Provides	Example
Surface form of term	glossSurfaceForm	The unambiguous full form of the term (which may include the acronym or short form in parentheses). The surface form is suitable to introduce the term in new contexts	<code><glossSurfaceForm> USB flash drive (UFD, or flashie) </glossSurfaceForm></code>
Part of speech	glossPartOfSpeech	The part of speech of the preferred form of the term.	<code><glossPartOfSpeech value="noun"/></code>
Symbol or icon	glossSymbol	Reference to an image file used as a synonym for the term.	<code><glossSymbol href="ufd_logo.jpg" scope="local"> <alt> Identification logo for USB flash drives </alt> </glossSymbol></code>
Notes on scope of usage	glossScopeNote	Information on what the term does or does not apply to.	<code><glossScopeNote>Not to be used for other flash memory cards that do not use a USB interface</glossScopeNote></code>
Status of term	glossStatus	Business rule recording the status of the use of the term. The status is usually explained with <code>glossUsage</code> .	<code><glossStatus value="obsolete"/></code>
Correct usage of term	glossUsage	Supplementary information explaining the correct use of the term.	<code><glossUsage>Do not use the term in title case (as in "USB Flash Drive") because that suggests a trademark.</glossUsage></code>
Variants of the term	glossAlt	Container for alternative representations of the glossary term (such as an acronym used for the term)	<code><glossAlt><... /></glossAlt></code>

In turn, the optional `glossAlt` element contains the following elements, used to store more detailed information about the glossary entry.

Element	Element Name	Provides	Example
Abbreviated term	glossAbbreviation	An abbreviated form of the term	<code><glossAbbreviation> Flash </glossAbbreviation></code>
Acronym	glossAcronym	An alternative form of the term as an acronym	<code><glossAcronym> UFD </glossAcronym></code>

Element	Element Name	Provides	Example
Association with another alternative form	glossAlternateFor	Cross-reference to a similar variant of the same term.	<glossAlternateFor href="#usbfd/memoryStick" />
Short form of term	glossShortForm	A short form of the preferred form of the term, but not an acronym	<glossShortForm> flashie </glossShortForm>
Status of variant	glossStatus	Business rule recording the status of the use of a variant of a term. The status is usually explained with glossUsage.	<glossStatus value="prohibited" />
Notes on usage of the term	glossUsage	Supplementary information explaining the usage of the variant of the term.	<glossUsage>This is too colloquial.</glossUsage>
Synonym	glossSynonym	A synonym of the glossary term.	<glossSynonym>memory stick</glossSynonym>

The glossgroup information type may be used to collect multiple glossary entries into a single topic. It serves a similar purpose to the *database* information type, being a nesting container for glossentry topics.

Example of a complete glossentry

The following is a glossary entry for USB flash drive.

```
<glossentry id="usbfd">
  <glossterm>USB flash drive</glossterm>
  <glossdef>A small portable drive.</glossdef>
  <glossBody>
    <glossSurfaceForm>USB flash drive (UFD, or flashie)</glossSurfaceForm>
    <glossPartOfSpeech value="noun" />
    <glossUsage>Do not provide in upper case (as in "USB Flash Drive") because that suggests a trademark.</glossUsage>
    <glossSymbol href="ufd_logo.jpg" scope="local">
      <alt>Identification logo for USB flash drives</alt>
    </glossSymbol>
    <glossScopeNote>Not to be used for other flash memory cards that do not use a USB interface.</glossScopeNote>
    <glossAlt>
      <glossAcronym>UFD</glossAcronym>
      <glossUsage>Explain the acronym on first occurrence.</glossUsage>
    </glossAlt>
    <glossAlt>
      <glossShortForm>flashie</glossShortForm>
      <glossUsage>Usually describes low cost devices.</glossUsage>
    </glossAlt>
    <glossAlt id="memoryStick">
      <glossSynonym>memory stick</glossSynonym>
      <glossUsage>This is a colloquial term.</glossUsage>
    </glossAlt>
    <glossAlt>
      <glossAbbreviation>stick</glossAbbreviation>
      <glossAlternateFor href="#usbfd/memoryStick" />
    </glossAlt>
    <glossAlt>
      <glossAbbreviation>flash</glossAbbreviation>
      <glossStatus value="prohibited" />
      <glossUsage>This short form is ambiguous.</glossUsage>
    </glossAlt>
  </glossBody>
</glossentry>
```

How a glossary entry might be rendered

Publishing tools may provide different methods of rendering glossary elements. The following shows an example of how a publishing tool may render a glossary entry in HTML form.

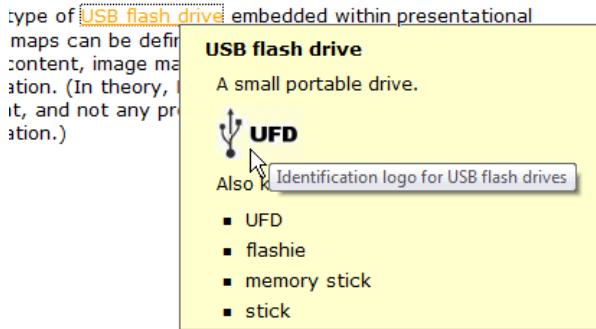


Figure 1: Hypothetical example of how a DITA glossary entry may be rendered in HTML.

Incorporating a glossary in a ditamap

Glossary topics (both glossentry and glossgroup) can be incorporated into a ditamap using a standard `topicref` element or through a new `glossref` element.

The `glossref` element has a required `keys` attribute used to help link terms in content topic (*inline terms*) to their glossary definitions. Rather than linking an inline term to a particular glossary topic file name, you can link the term to the `keys` attribute of the relevant `glossref` entry. This *indirect addressing* mechanism allows greater flexibility for single-sourcing, where the glossary definition might be different for different collections of topics.

The diagram below shows the `glossref` being used for indirect linking of terms.

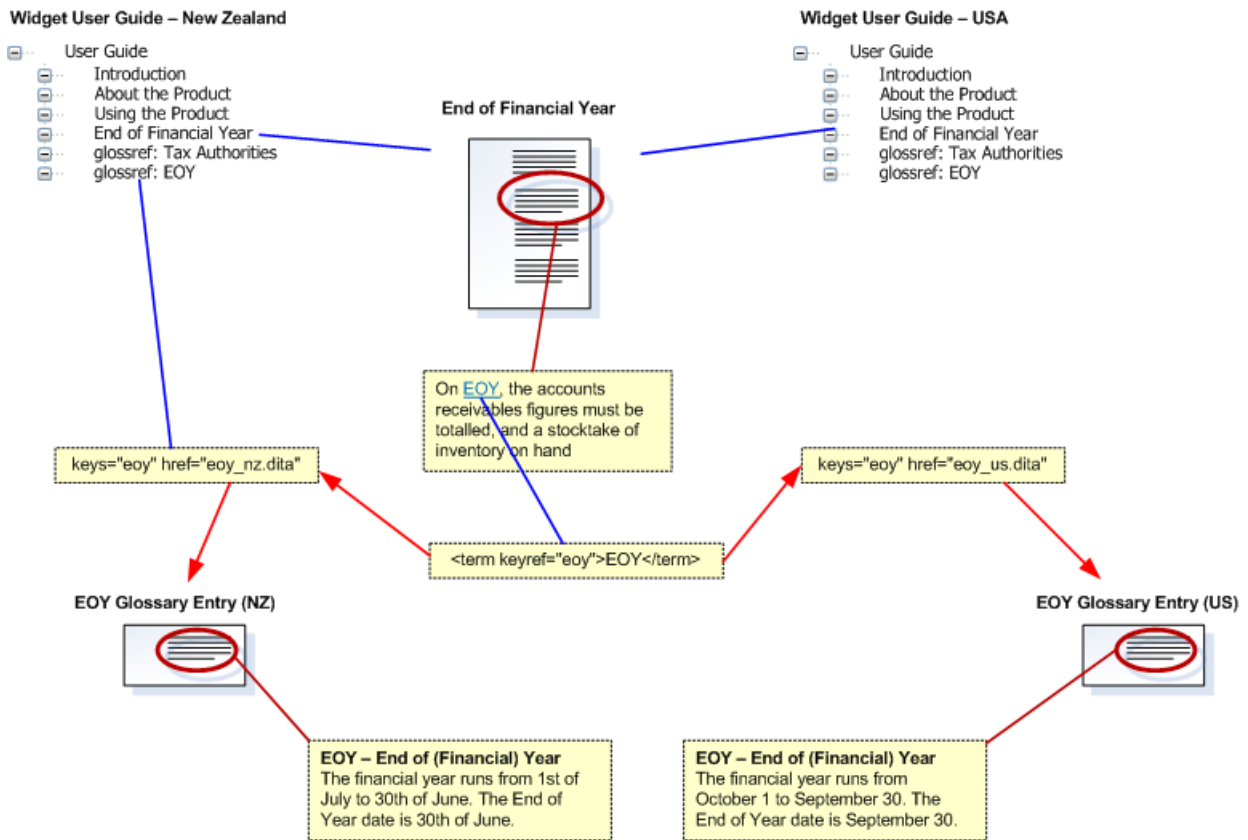


Figure 2: Indirect Referencing of Glossary Topics using keys

An important difference between `glossref` and `topicref` is that `glossref` won't result in the glossary topic being inserted in the output at the position the reference appears in the ditamap.

Creating a glossary section

A glossary section in an output document is defined by referencing, through normal topicref elements, all the glossary (glossentry and glossgroup) topics in a normal ditamap. The topics must be manually sorted into alphabetical order unless the authoring tool offers an automated topic sorting feature.

```
<map>
.
<topicref href="abs.dita" />
<topicref href="awd.dita" />
.
<topicref href="wrx.dita" />
.
</map>
```

Linking terms in content topics

The linking of a term occurring inline in a content topic to its relevant glossary definition is a matter of associating the term (semantically identified in a term element) with the hypertext reference (href) or the key reference (keyref) of the glossary topic.

For example, the markup `<term keyref="eoy">EOY</term>` establishes the association between the term “EOY” and the glossary topic with a keys attribute of eoy. The related map entry may be `<glossref keys="eoy" href="eoy_nz.dita"/>`.

In addition to using the term element as the basis for linking, a new abbreviated-form element provides an alternative method of using glossary entries by substituting the element for the glossSurfaceForm or glossAcronym variants from a glossary entry on rendering. It is expected that a processor will substitute the first occurrence of an abbreviated-form element with the glossSurfaceForm variant in the glossary entry and subsequent occurrences with the glossAcronym variant of the glossary entry. The abbreviated-form element acts like a content reference but uses the key rather than the href as the reference.

For example, markup of:

```
<p>An <abbreviated-form keys="abs"/> helps a driver to stop. For
this reason many find an <abbreviated-form keys="abs"/> useful.</p>
```

might be rendered in HTML as:

```
<p>An Anti-lock Braking System (ABS) helps a driver to stop. For
this reason many find an ABS useful.</p>
```

where the glossref in the map was `<glossref keys="abs" href="abs.dita"/>`, and the abs.dita glossary entry topic was:

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

About keys and keyref attributes, and indirect linking

Whenever a topic has a reference to other content, it makes the topic less reusable because of the dependency on the target being still available and still relevant. The keys and keyref attributes provide a simple redirection scheme that leverages existing attributes and map architectures to provide support for redirectable conrefs, topicrefs, xrefs, links, terms, and other reference elements and attributes. They also provide a simplified architecture for managing variable or volatile content (such as product names) which need to be easily swapped out when a topic is reused in new contexts.

A topic's keys are always defined in the map. A keys attribute can contain one or more keys. Keys resolve to the resources given as the href value on the topicref element with the matching key.

Summary

The new glossary features in DITA 1.2 offer many opportunities for processing glossaries and terms into rich features in output documents. The enhancements in DITA include new glossentry and glossgroup information types, a new glossref map element, a new abbreviated-form alternative to the term element, and the addition of a keys attribute to the term element to allow for indirect linking.