

Sonata Software Limited, Microsoft Corporation and DAISY Consortium

# **AUTHORING GUIDELINES FOR OPEN XML TO DAISY XML TRANSLATOR**

Authoring Guidelines

# Authoring Guidelines for Open XML to DAISY XML Translator

---

## Contents

1	Introduction .....	3
2	Features .....	4
2.1.1	UID Metadata [R2] .....	4
2.1.2	TABLES [R7] .....	5
2.1.3	Images [R14] .....	5
2.1.4	Captions [R18].....	6
2.1.5	Notes and Notes References [R10] .....	8
2.1.6	Paragraphs [R5].....	8
2.1.7	Lists [R8] .....	9
2.1.8	Levels and Headings [R4] .....	9
2.1.9	Block quotes.....	10
2.1.10	Abbreviations and Acronyms.....	10
2.1.11	Page Numbers.....	10
2.1.12	Side Bar .....	11
2.1.13	Table of contents .....	13
2.1.14	Track Changes .....	14
2.1.15	Custom styles .....	14
2.1.16	Math ML.....	16
2.1.17	Multiple OOXML.....	16
2.1.18	Language Detection .....	18
2.1.19	Shapes And Objects.....	18
2.1.20	Full Audio Support.....	19

## 1 INTRODUCTION

This document provides guidance on how to create an Office Word document to minimize the fidelity losses while translating to DAISY xml.

## 2 FEATURES

Following features will need to be used as per the guidelines given in this document for minimum fidelity losses:

- UID metadata
- Tables
- Images and Image group
- Captions
- XML Data
- Footnotes & Endnotes
- Paragraph
- Lists
- Headings
- Page Numbers
- Block Quotes
- Abbreviation and Acronyms.
- Side Bars
- Table of content
- Custom styles
- MathML (Support for Mathematics)
- Multiple OOXML
- Language Detection
- Shapes And Objects
- Full Audio Support

### 2.1.1 UID METADATA [R2]

- An auto generated value will be displayed in a textbox “Document UID” on the user interface; this ID uniquely identifies a document. User can change this value if required.
- If user deletes this value by leaving the field Document UID empty, an auto generated ID will be populated with a prefix “AUTO-UID” and a randomly generated number.

## 2.1.2 TABLES [R7]

- Complex tables in DAISY are termed as long tables crossing several pages of the print book. Whereas, in Open XML the complex tables includes table having rows/columns merged, table inside a table, table with background color and borders with different shading colors, tables with page-bordering, tables with text color, table having images.
- The mapping of these complex tables from Open XML to DAISY is discouraged as it may lead to fidelity losses. Users are encouraged to use simple tables with rows and columns.
- Simple tables with rows and columns, large tables spanning across one or more pages are implemented with minimum fidelity losses.
- For a Table it is advisable to use Table level cells alignment.
- In a scenario where a table is inserted in between the text, then tables will be displayed before the text.
- It is highly recommended to use tables for tabular data and not for layout purposes. It is better to use styles specifically for the formatting you want to accomplish.
- In the translation process, headings and sidebars etc. found inside a table will be converted to paragraphs. If you want to keep these as headings or side bars, please move them outside the table and use the appropriate style.

## 2.1.3 IMAGES [R14]

- Simple images like .jpeg, .bmp, and .gif will be converted without any fidelity losses. However, users are highly recommended to use alternate text for image/image group. If a user is not entering alternate text for images then the output xml file will contain alternate text attribute without any value in it.
- Complex images like image embedded in any shape object, image in a table, different pictorial shapes like polygons, and clipart if implemented may lead to fidelity losses.
- Image groups should be implemented using a canvas. Ungrouped images inside the canvas may lead to fidelity loss.
- In the case of embedded objects such as images, object will be copied to the destination folder and the relationship will be set accordingly.
- Image if inserted in between the Text then Image will be displayed before the text.
- Image files with “.jpeg”, “.jpe” and “.jpg” extensions will all be converted to files with “.jpg” extension and copied to the destination folder.
- All other image files will be converted to files with “.png” extension and copied to the destination folder.
- File names of images copied and pasted into the word document will not be retained in the output xml and in the destination folder.
- Images which are imported from external resources will not be translated.

# Authoring Guidelines for Open XML to DAISY XML Translator

---

- Caption provided above the image will not be translated. For proper translation captions for images should be provided below the image.

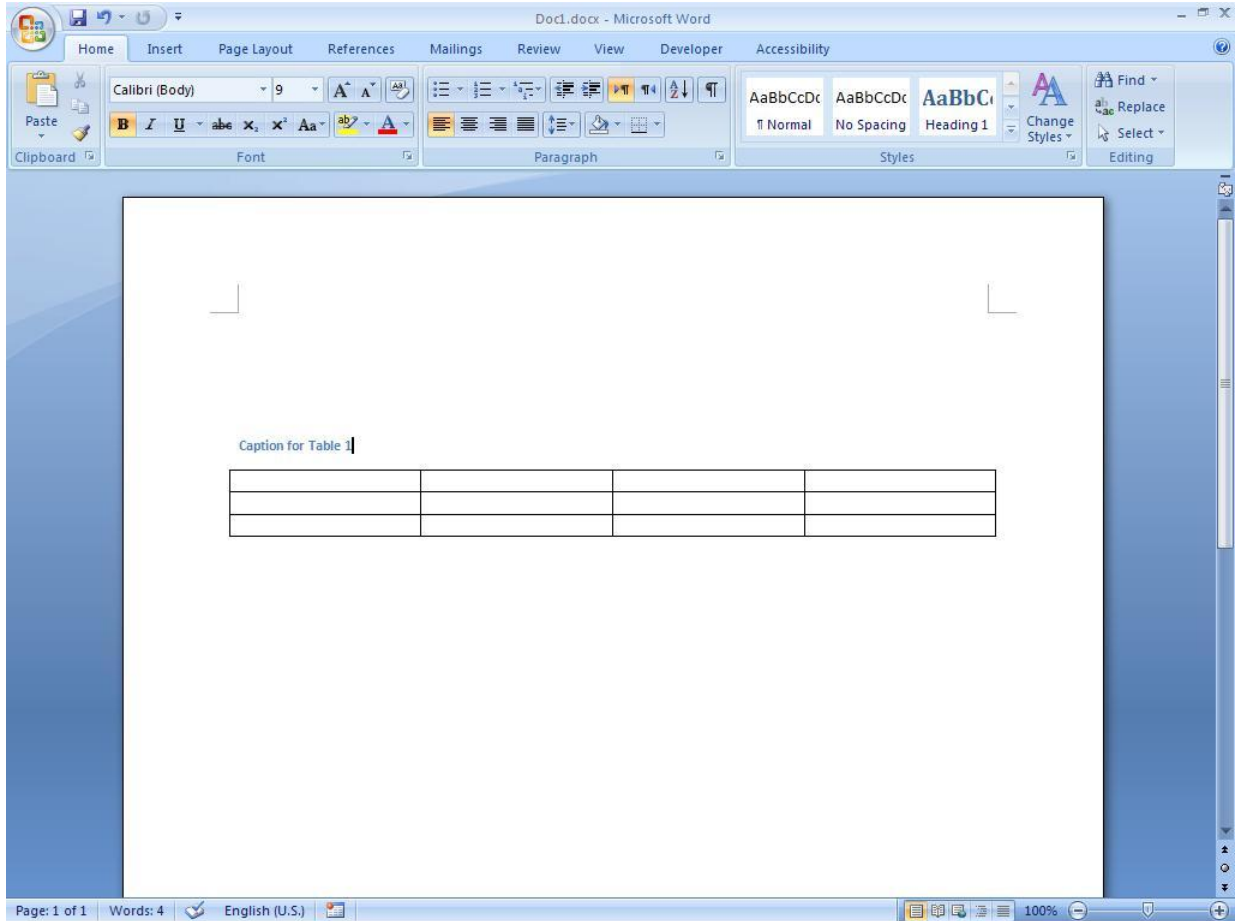
## 2.1.4 CAPTIONS [R18]

- The captions in DAISY are tightly coupled with their respective tables/image group. In Open XML the captions for simple images/image group/tables are loosely coupled.
- The users are highly recommended to use the captions below the image(s) for image/image group and for the table the caption should be above the table for minimum fidelity losses.
- If images wrapped with text using other than “In line with Text” option, Captions will not be handled.

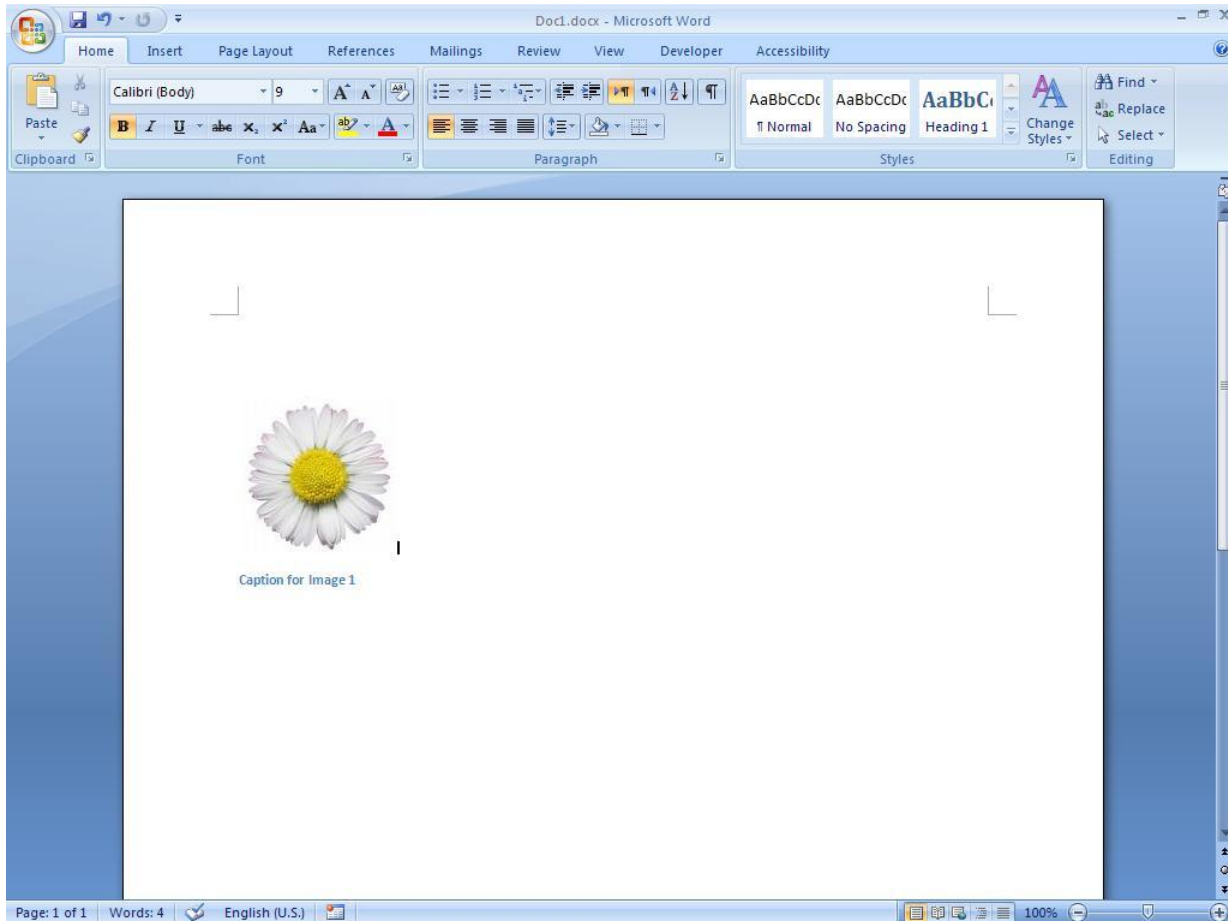
- **Caption for Grouped Images:**

Users are recommended to use “Image-Caption (DAISY)” custom style for grouped images instead of inbuilt captions.

# Authoring Guidelines for Open XML to DAISY XML Translator



# Authoring Guidelines for Open XML to DAISY XML Translator



## 2.1.5 NOTES AND NOTES REFERENCES [R10]

- Footnotes will be populated at the end of the page. If the footnote is not populated at the end of the page, it will be populated in the end of the next page otherwise it will be populated at the end of the document.
- If footnotes containing images and shapes, Images and shapes data will be lost.

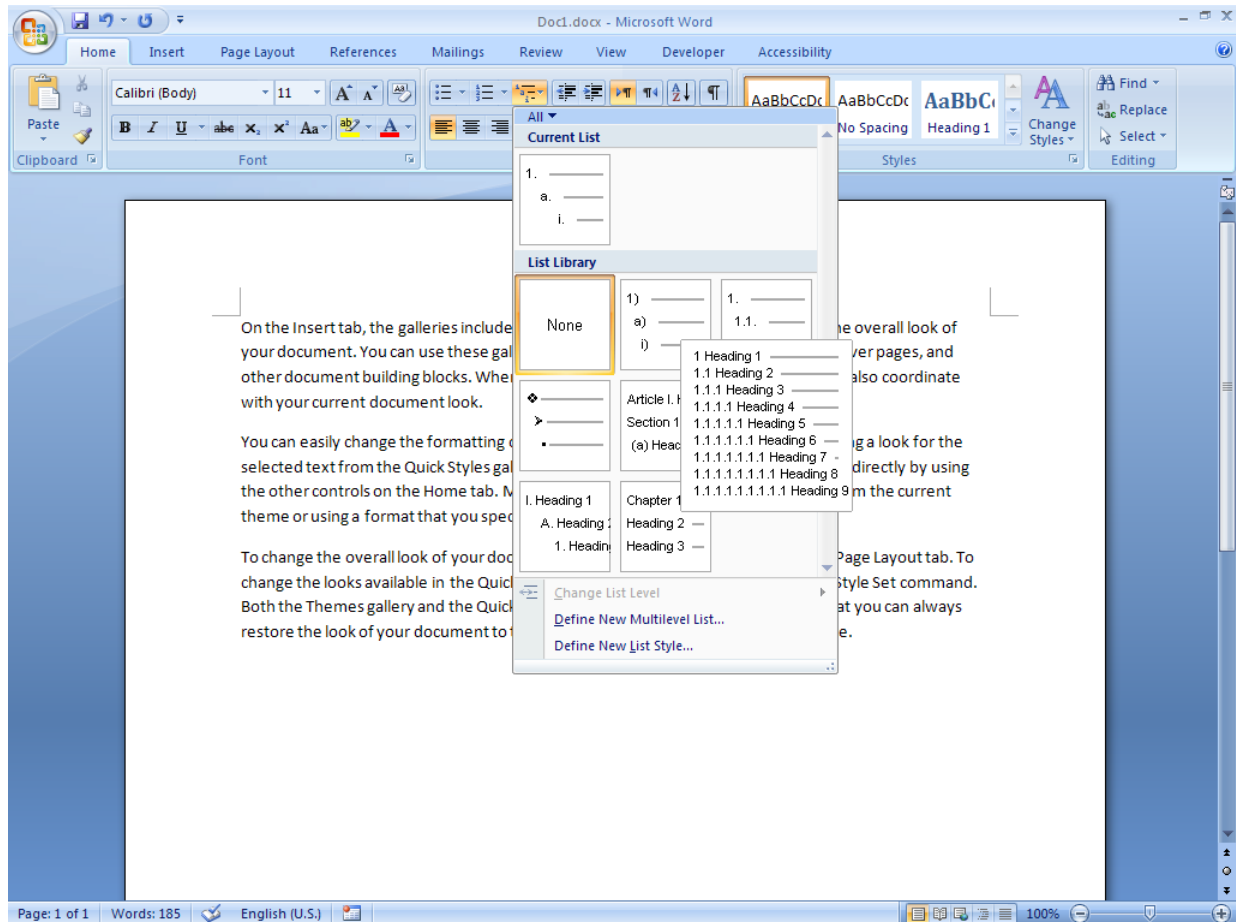
## 2.1.6 PARAGRAPHS [R5]

Styles like Intense Emphasis and Subtle Emphasis are not handled as it may lead to fidelity loss. For features like Emphasis and Strong, paragraph styles *Emphasis* and *Strong* should be used. However, users can also use Bold and Italic text styles for Strong and Emphasis features respectively.

# Authoring Guidelines for Open XML to DAISY XML Translator

## 2.1.7 LISTS [R8]

- Numbered list will be converted as an ordered list.
- It's recommended not to insert List of Bridgehead in a word document.
- Only the default list styles will be supported. New (Custom) list styles will not be supported.



## 2.1.8 LEVELS AND HEADINGS [R4]

- In a scenario where heading appears between running text then the translated sentence will be broken with heading information appearing as simple text.

---

# Authoring Guidelines for Open XML to DAISY XML Translator

---

## 2.1.9 BLOCK QUOTES

- Tables should not be inserted inside the Block quotes

## 2.1.10 ABBREVIATIONS AND ACRONYMS

- Abbreviations and Acronyms are handled with Bookmark feature of the Word. All words marked as Abbreviation and Acronyms are internally marked as Bookmark with unique name which starts with "Abbreviations" and "Acronym" respectively. So user will see many bookmarks in the Bookmark dialog box of word, where the name of the bookmark starts with "Abbreviations" and "Acronym". User should not delete these bookmarks. If user deletes the bookmark, they will be losing the Abbreviations and Acronyms information.
- User should not use the name which starts with "Abbreviations" and "Acronym" for their bookmark.
- In Word 2003 and Word XP, application uses custom document properties to store the expansion of Abbreviations and Acronym. So user should not delete the custom document properties, where the name of the properties starts with "Abbreviations" and "Acronym"
- In Word 2003 and Word XP, User should not use the name which starts with "Abbreviations" and "Acronym" for their custom document properties.

## 2.1.11 PAGE NUMBERS

- User should not span a table across multiple pages without any text in the cell.
- User should not start a new page with a blank line.
- User should not start a new page with an image.
- Table of content spanning multiple pages will be translated to one single page. For Example if the Table of contents is spanning across page 3 and 4, it will be considered as the entire Table of content in the page 3.
- In multi column documents, each column in the page will be considered as a single page.
- Page number with chapter number is not supported, so user should not format the page number with chapter numbers.
- Formatting the page numbers by editing the field is not supported. User should use Page number format option for formatting the page numbers.
- Only Simple type of page number from the gallery of Page number designs will be supported. Use Simple type of Page number design.
- Page numbers are handled in the best possible way, still there may be some scenarios will not result in proper page numbers. For example

# Authoring Guidelines for Open XML to DAISY XML Translator

---

- Some documents with empty pages (based on the user behavior) will not result in proper page numbers.
- If the document is not having the proper page break information, it will result in improper page numbers.
- It is highly recommended to use custom style for implementing page numbers, as the automatic page numbers may have some issues.

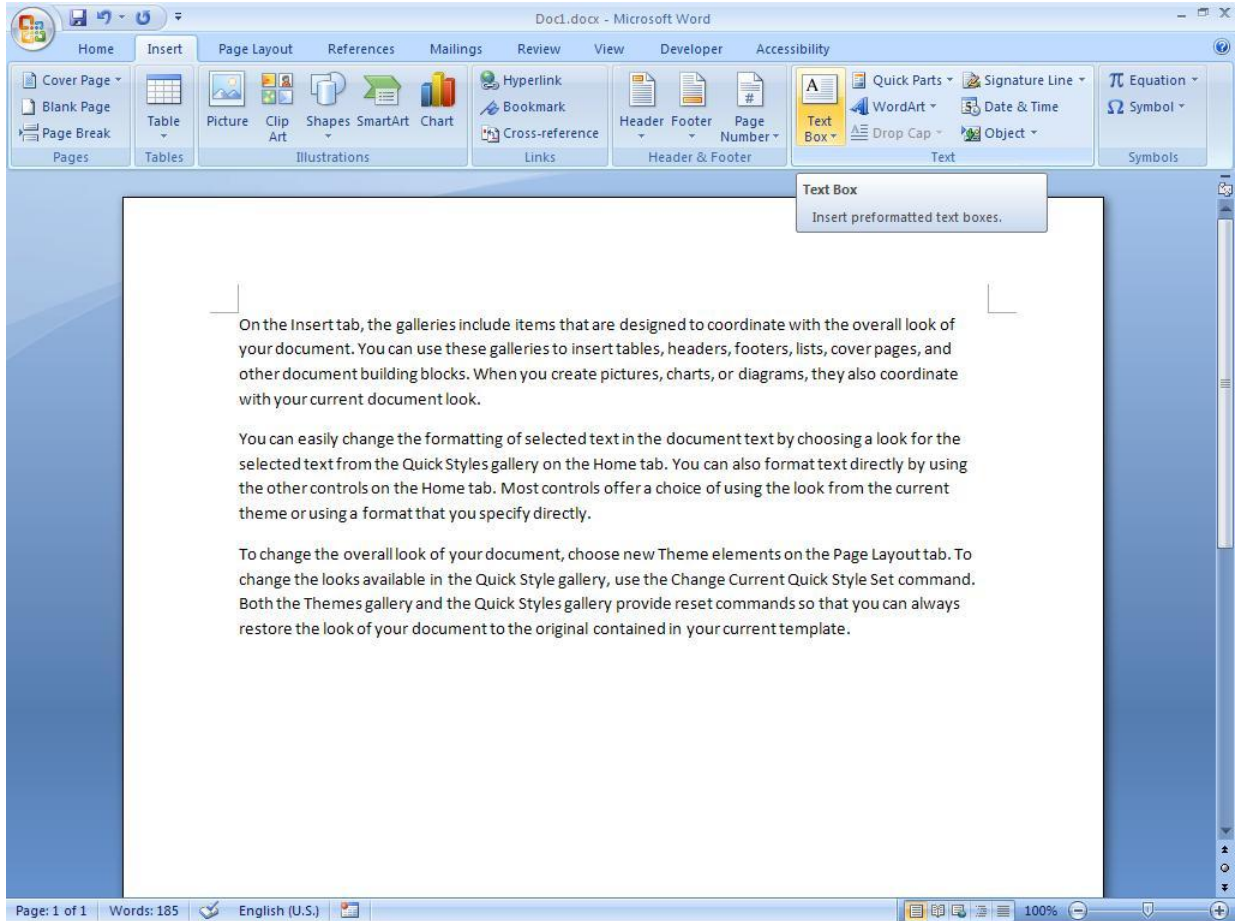
## 2.1.12 SIDE BAR

User should follow the following steps to insert a sidebar to a word document.

The steps are explained with the help of screenshots.

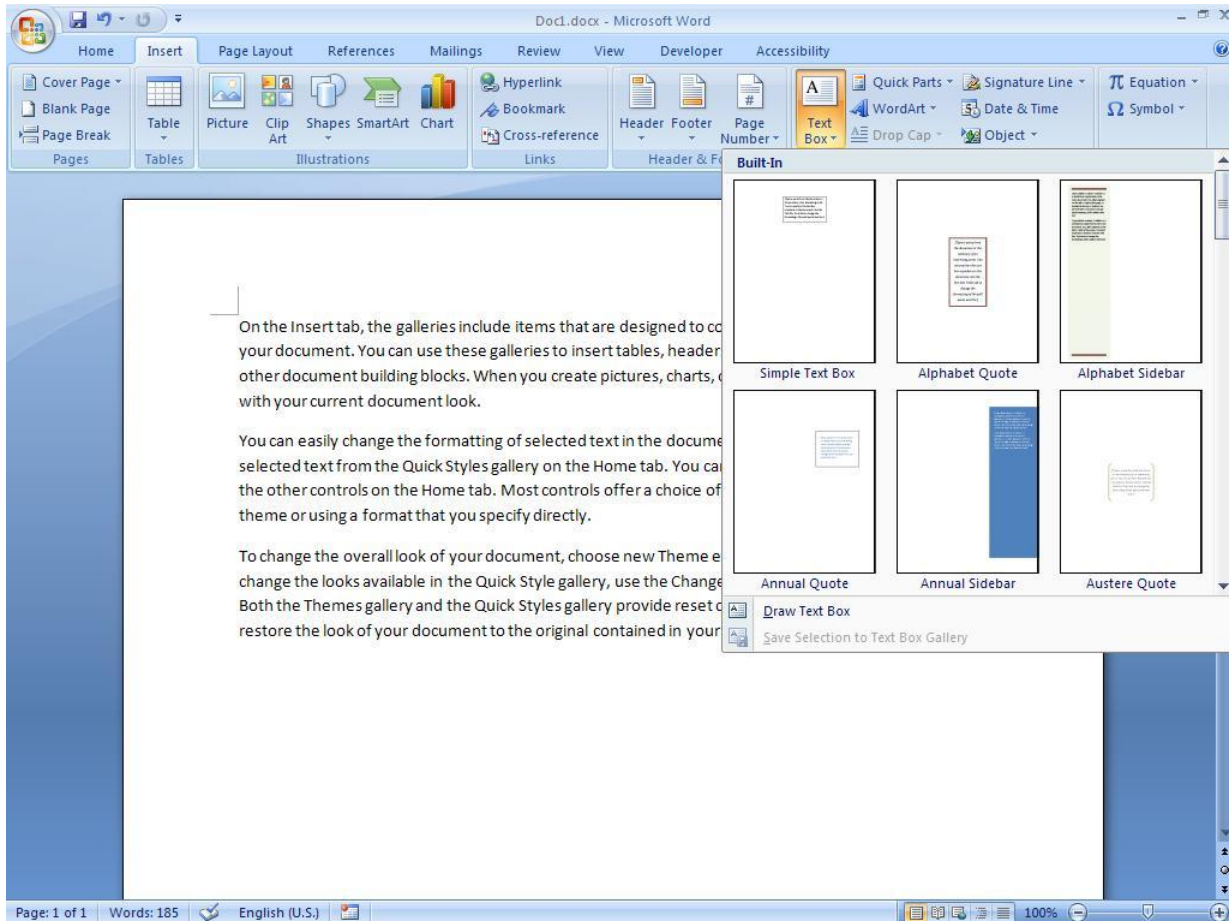
- Below screenshot shows the options of Textbox highlighted after clicking on Insert tab in Tool bar.
- Images and tables with caption inside a sidebar may lead to some information lost. So, it is highly recommended to use custom style of caption for images and tables inside a sidebar.

# Authoring Guidelines for Open XML to DAISY XML Translator



- After clicking on Textbox option a user will get the various types of Sidebars which are available.
- User can click on sidebar option and can insert it in word document.

# Authoring Guidelines for Open XML to DAISY XML Translator



## 2.1.13 TABLE OF CONTENTS

- Word provides three types of built-in styles for Table of content. Manual Table of content is one of the styles. User should not use this option. User should use Automatic styles of Table of contents.
- Table of contents should not be embedded with other features of Word, like sidebars, Table cell etc.
- Table of content is always expected to be at the first page of the word document. In case it is in the middle of word document, even then the converted Xml will have it in First page.
- After inserting the Table of Content user should not change the contents. If they change the content, it will produce unpredictable result.

## 2.1.14 TRACK CHANGES

Before translating the word document to Daisy document ensure that

- “Track Changes” is OFF for a document.
- All changes are either accepted or rejected.

### Comments:

Comments in a document may not get translated. So before translating the word document to Daisy document ensures that document does not have comments.

## 2.1.15 CUSTOM STYLES

Image caption should be inserted at the Top or Bottom of an Image, if it is provided at both the places then the resultant caption will contain both the captions concatenated.

### 1. Prodnote for Images (Prodnote):

- Producer Note for images should always be inserted below the image without empty spaces.
- Prodnote for images inside the block quote will not be translated.
- Producer note inserted above the Image will not be captured under <imggroup> Tag.

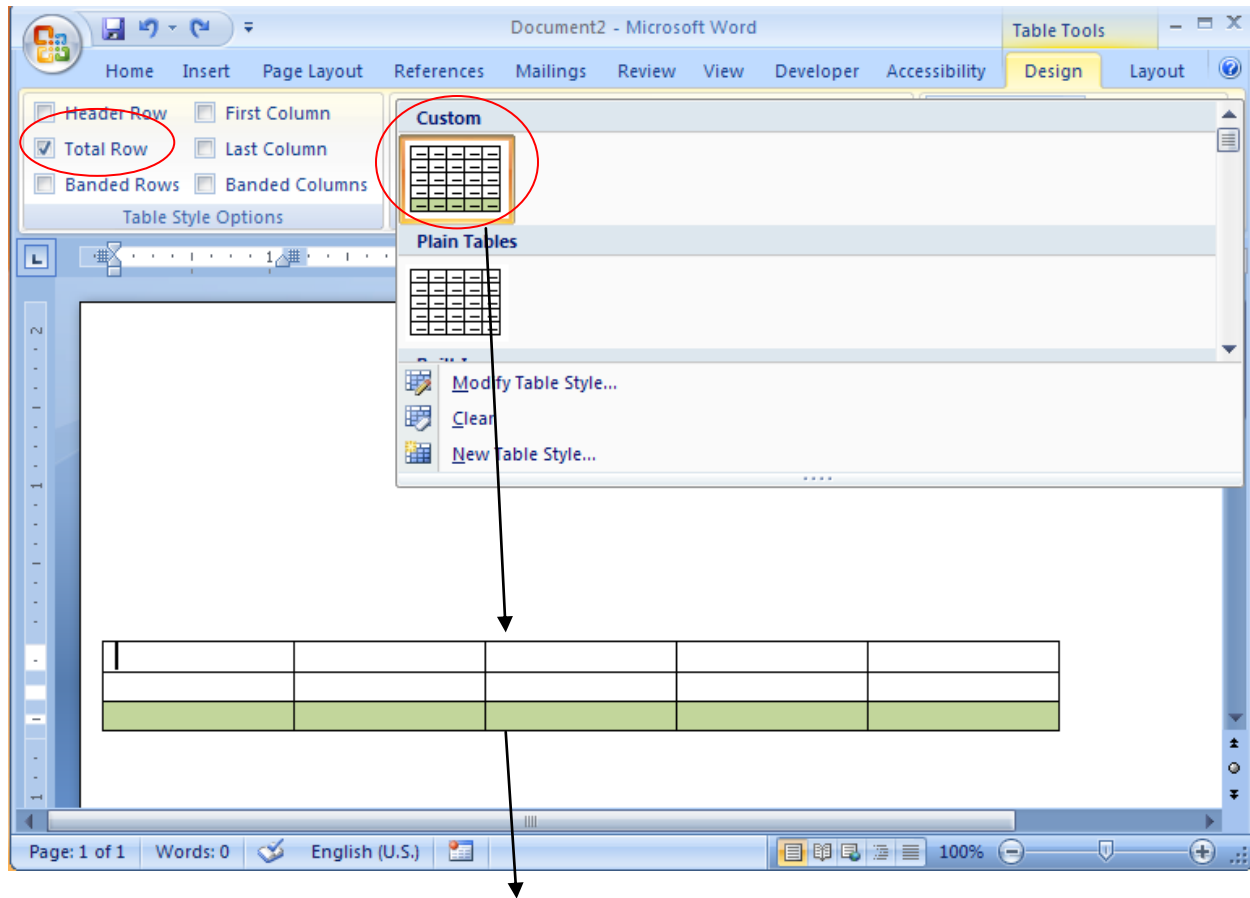
### 2. LineNumbers:

- Line numbers for a bdo (bidirectional oriented) paragraphs will not be translated.
- If line number custom style is applied for the first line of a paragraph, then empty <p> tags will occur in the output xml file.

### 3. Table Footer:

- Table-Footer (DAISY) custom style is applied always to the last row of the table. This custom style for table will automatically be updated in the Table Tools tab on click of “Import Daisy Styles” button of the Accessibility tab as shown in the screen-shot below:

# Authoring Guidelines for Open XML to DAISY XML Translator



**Table-Footer (DAISY)**

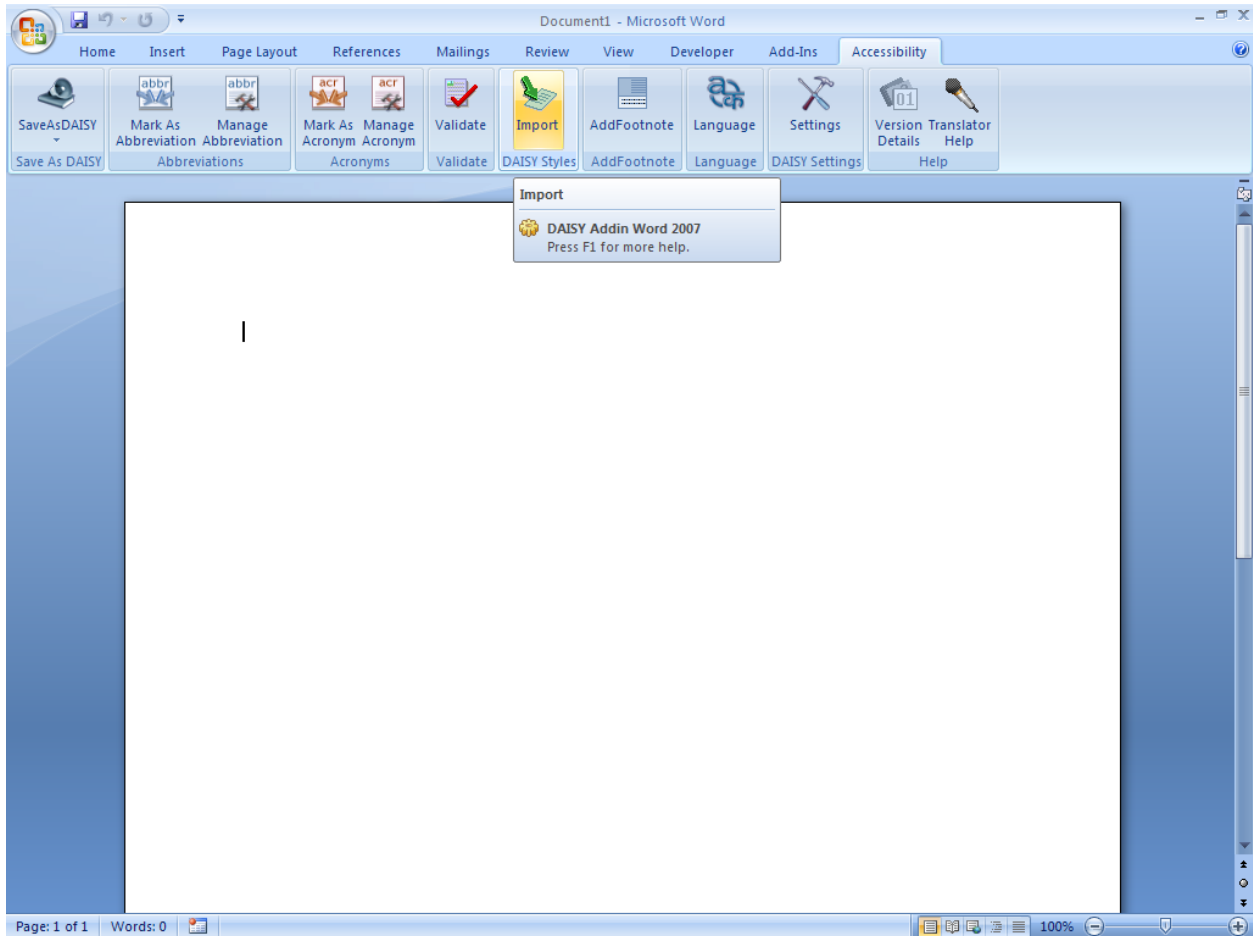
To apply this style, user has to select a required table and click on the **Table-Footer (DAISY)** style and also click on the “**Table Row**” check box as shown the screen-shot above.

## Import Daisy Styles:

If the default styles (i.e. Default Paragraph Font, Normal, Table Normal, No List) of the document has been changed, when importing the Daisy styles from the template to the active document, all the default styles which are changed has been revert back to the original style. This makes the document structure to be changed.

- To Import all the Daisy Styles from the template document to the Active document, “Import Daisy Styles” button need to be clicked.
- There are around 35 daisy styles which need to be imported to the active document. When all the styles are imported, “Import Daisy Styles” button should be disabled.
- If a new style has been added or any style has been deleted in the template, and if the document which contains only the 35 daisy styles is opened, the button “Import Daisy Styles” needs to be again enabled.

# Authoring Guidelines for Open XML to DAISY XML Translator



## 2.1.16 MATH ML

- Equations in Word 2003/XP will be converted as Images (PNG Images).
- Reference to footnote should be made outside the equation block, as marking it inside marks the reference also as a Math Equation.
- Language support for MathML will not be provided.
- Headings and Styles should not be applied to Math Equations.
- In Command Line Utility, if Math Equations inserted through MathType software, then equations will not be translated.

## 2.1.17 MULTIPLE OOXML

- Table of Contents of Sub documents will not be translated.
- Cover pages of sub documents will not be translated.

# Authoring Guidelines for Open XML to DAISY XML Translator

---

- Master Document should not have Anchor to a subdocument.
- Cover pages for both Master document and Subdocument will not be translated.
- Master document having subdocument should not be selected for Translating set of docs.
- While translating the set of documents, all the documents should be closed.
- Metadata information (Title, Creator etc) of sub documents will not be translated.

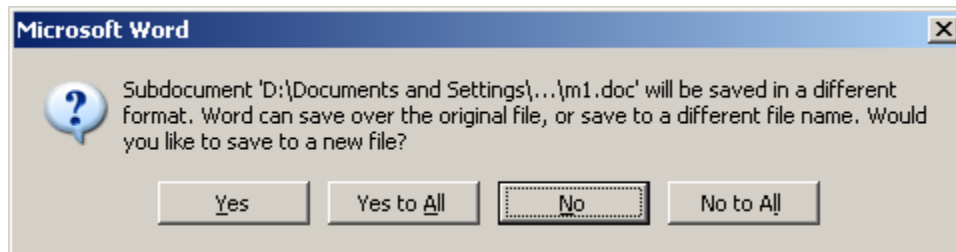
User should follow the following steps to create a master/sub document in respective environments.

## Word XP:

Creating master document in Word XP:

1. Create a master/sub document in Word XP (.doc) format.
2. Save master/sub document in Word 2007(.docx) format.

While saving .doc file to .docx file, on click of “Save as” button the following window will be prompted to the user.

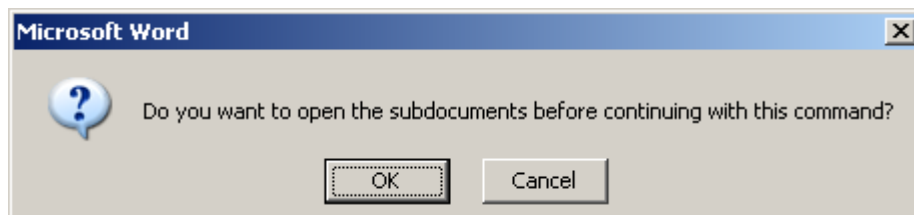


Users are advised to choose “No” option so that the document is saved in word 2007 format, but the extension of the document will be displayed as “.doc” in word editor, which should not be concerned..

## In Word 2003 and Word XP:

Opening a Word 2003/XP document all sub documents will be closed (appear as Hyperlink) in word editor.

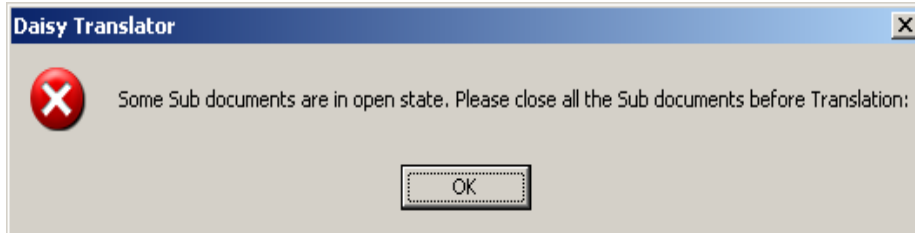
On clicking “Save as Daisy” button; the following window will be displayed to the user to open the subdocuments, User is advised to close all the documents opened before Translation.



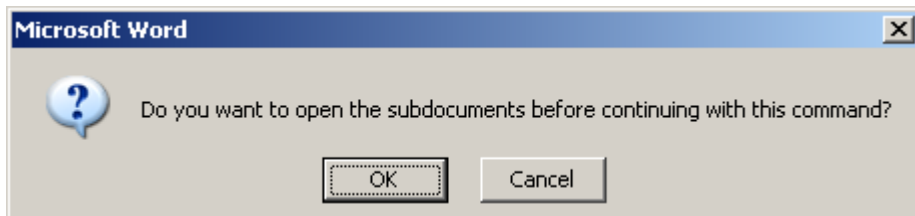
# Authoring Guidelines for Open XML to DAISY XML Translator

---

If any document is left opened (Expanded hyperlink) following window will be prompted indicating some word document is open, close the document and proceed.



Again Click on "Save As DAISY" button the same window will appear even if all the documents are open, it is advised to ignore the message and click on "OK" button.



Clicking on "OK" Button the translation process will be initiated and all the subdocuments will be converted to single DTBOOK.xml.

**Note:** While translating the master document, all the sub documents should be in closed format (Hyperlinks Unexpanded).

## 2.1.18 LANGUAGE DETECTION

Language feature works only at paragraph level. If user selects a part of a paragraph, the language information will not get translated properly. So the translation would be effective if user selects the entire paragraph or more than one paragraph.

## 2.1.19 SHAPES AND OBJECTS

- In Command Line Utility, Shapes and objects will not be translated.
- Audio Books for a document containing embedded objects will not be created. Example, a word document is embedded inside a word document. Users are advised to create audio books separately for the embedded documents.

# Authoring Guidelines for Open XML to DAISY XML Translator

---

## 2.1.20 FULL AUDIO SUPPORT

- In Command Line Utility, option for generating audio files is not supported.
- To generate the audio files from DAISY Xml through Command Line, user should use Pipeline command line utility and it is available in Application Folder of Daisy Translator. For example, if you have installed DAISY Translator in “C:\Program Files\Sonata\Daisy Translator Addin Version 1.5.0.0”, then the Pipeline Command Line utility named as “pipeline-lite.bat” will be located in “C:\Program Files\Sonata\Daisy Translator Addin Version 1.5.0.0\pipeline-lite-ms”.