# OSIS to eBook Conversion

## Introduction

The OSIS to eBook conversion tool converts OSIS-encoded Bibles to various eBook formats such as EPUB, MOBI and FB2 (Fiction Book version 2). It is intended to be used with OSIS files created from paratext by usfm2osis.py , and may not work satisfactorily with OSIS files from other sources. In the future, the tool may be extended to work with other OSIS documents, including commentaries and genbooks.

The tool consists of:

* The Perl script osis2ebook.pl
* The Python script fb2postproc.py
* The cascading stylesheet ebible.css
* The Calibre plugin osis-input, which is a collection of Python scripts.

Future releases of the tool may include additional scripts and cascading stylesheets.

## Prerequisites

### 2.1 Perl

Perl is required for the conversion tool as well as for osis-converters. Perl is likely to already be available on Linux systems.

### 2.2 Python

Python is required for the conversion tool. The recommended version of Python to use is 2.7. Python 3 should not be used. Python is likely to already be available on Linux systems. Note that if you have multiple versions of Python installed on Linux, it may be necessary to edit the initial line of fb2postproc.py to ensure that the correct version is used.

### 2.3 Calibre

This tool requires version 2.15 or later of Calibre. The installer can be downloaded from:   
 <http://calibre-ebook.com/download>.

Once Calibre has been installed, the osis-input plugin must be set up. If you have the plugin as a zip file, the required command line is:

calibre-customize –a <zip file path>

If you have the plugin in uncompressed form, the required command line is:

calibre-customize –b <path of directory containing plugin>

## Preparing to Run Conversion

The conversion tool requires the following files in a single directory:

* The OSIS file to be converted.
* OSIS files for any associated glossaries.
* A text file named convert.txt which contains information required for the conversion.
* The cascading stylesheet file ebible.css

If there are any images to be included (referenced from <figure> tags in the OSIS file), these must be in a subdirectory of this directory, named ‘images’. The only supported image file types are JPEG (file extension .jpg or .jpeg), GIF (file extension .gif) or PNG (file extension .png).

If ebible.css references fonts which are to be embedded in the eBook, it should be in a ‘css’ subdirectory along with the required font files. Fonts may be either OpenType fonts (file extension .opf) or TrueType fonts (file extension .ttf).

### 3.1 OSIS File

The OSIS file must have the extension .xml, since the osis-input plugin registers itself to handle files with this extension. It should have been created using the current version of osis-converters. Converting files created with older versions may not produce satisfactory results.

A glossary OSIS file must also have the extension .xml. The file name must match the name of a glossary referenced from the main OSIS file.

The ordering of the books of the Bible in the eBook will be the same as the ordering in the OSIS file.

The conversion does not include scripture reference links or cross-references. If these are included in the OSIS file they will be ignored, but they could all be excluded by specifying in CF\_usfm2osis.txt:

SET\_addScripRefLinks:false

SET\_addCrossRefs:false

### 3.2 convert.txt

The file convert.txt contains information required for the conversion process. Each line is of the form:

<property>=<value>

The file may also contain blank lines and comment lines beginning with #.

The file must be UTF-8 encoded, and is expected to have no BOM (byte order mark). If using an editor which automatically includes BOM, such as Notepad on Windows, ensure that the first line is blank or a comment line, as option values on this line will not be recognised.

#### Book Titles

Titles of the books of the Bible may be specified as a number of property-value pairs. This is necessary where the book titles are not present in the OSIS file. The property to the left of the equals sign is a book of the Bible as specified in OSIS IDs, e.g. “Gen” or “Matt”. The value to the right of the equals sign is the book title as it is to appear in the eBook, e.g.

Gen=ИБТИДО  
Exod=ЧИҚИШ  
…

Some OSIS Bibles include book titles as headings in the book introduction or as initial section headings. Where this is the case, book titles may be extracted in from the OSIS file by including the line:

BookTitlesInOSIS=True

Where this is specified, a title at that start of a book is assumed to be the book title. If titles are included in the OSIS file only for some books, titles for the other books may still be specified in convert.txt, but book titles extracted from the OSIS file take priority.

There are some cases where it may be necessary to specify book titles in convert.txt even though they are present in the OSIS file. This will be the case, for example, where a Bible or Testament introduction precedes the book title for the first book of the Bible or Testament. The titles specified in convert.txt should match exactly the titles in the OSIS file; the conversion tool will detect that the titles match and so will recognise the book title even where it does not occur at the start of the book.

If there is any case where a book starts with a title which in not the book title, “BookTitlesInOSIS=True” should not be specified, as in the case where the book title is preceded by a Bible or Testament introduction which has a title. If titles in convert.txt match exactly the book titles in the OSIS file, the conversion tool ensures that duplicate titles do not appear.

#### Group/Testament Titles

The properties Group1 and Group2 can be used to provide titles for book groups in the OSIS file, where the book groups are actually the Old and New Testaments, e.g.

Group1=Old Testament  
Group2=New Testament

The titles specified will appear on title pages before the first book of each Testament, and will also appear at the top level of the table of contents. For an OSIS file containing only New Testament books, these properties should not normally be specified – if they were, only the Group1 title would be used. If these properties are not specified when the OSIS file contains both Old and New Testament books, there will be no division between Old and New Testament books in eBook or its table of contents.

If testaments are not in separate book groups, the Group1 and Group2 attributes can still be used to provide testament titles by including the line:

TestamentGroups=False

#### Metadata

The book title, publisher and language metadata items can be specified in convert.txt, e.g.

Language=bua  
Publisher=Институт перевода Библии  
Title=Шэнэ Хэлсээн

If these are not specified in convert.txt, publisher is set to “IBT” and language and title are extracted from the OSIS header. In this case the title will be “<xxx> Bible”, where <xxx> is the module name, e.g. BUR or UZV, which is unlikely to be what is wanted. There is currently no way to specify other metadata items such as author or ISBN. The cover image metadata is set up from the command line, as will be described later.

#### Chapter and Psalm Titles

Chapters and psalms may either start with a large chapter or psalm number on the left of the first verse – a practice adopted in many editions of the Bible – or a heading such as “Chapter 1” or “Psalm 23”. Some OSIS Bibles specify chapter/psalm headings. Where this is not the case, and headings are wanted, then the form that these take is specified using the ChapterTitle and PsalmTitle properties. The corresponding value for each attribute is a template containing “%s” which will be replaced by the chapter or psalm number, e.g.

ChapterTitle=%s–БОБ  
PsalmTitle=%s–САНО

The PsalmTitle template is used in the book of Psalms, and the ChapterTitle template in all other books. So with the above values, the heading for Psalm 1 would be “1–САНО” and the heading for chapter 2 in any other book would be “2–БОБ”.

Where these properties are not specified (or commented out) and no headings are specified in the OSIS file the large chapter/psalm number format is used for EPUB and MOBI output. It is acceptable to provide only one of these properties. If, for example, PsalmTitle is specified but not ChapterTitle, psalms will have title headings, but chapters in all other books will have large chapter numbers.

The large chapter/psalm number format is not possible for FB2. If ChapterTitle and PsalmTitle templates are not provided and no headings are specified in the OSIS file, the chapters and psalms will have headings consisting of just the chapter/psalm number.

Titles specified in the OSIS file take precedence over ChapterTitle and PsalmTitle templates.

#### Psalm Division Titles

In many Bibles, the book of Psalms is divided into 5 parts, with a heading for each part. The conversion tool needs to be able to recognise such headings, so that they are correctly positioned and formatted. The property PsalmDivTitle provides a regular expression to match such a heading. Any heading in the book of Psalms which matches this expression is treated as the title of one of these divisions. Typically, such a heading is followed by a sub-heading indicating the range of psalm numbers. The property PsalmDivSubTitle provides a regular expression to match such a sub-heading. These sub-headings are recognised by the conversion tool only if they directly follow a division heading. Here is an example of the specification of these properties:

PsalmDivTitle= Китоб \([1-5]\)  
PsalmDivSubtitle=\([0-9]+-[0-9]+–Санолар\)

The above would, for example, match the heading “Китоб (2)” followed by the sub-heading “(41-71–Санолар)”.

Without these properties specified, these headings would be formatted and positioned as normal section headings, which means that they would appear after rather than before the heading for the following psalm.

#### Bible and Testament Introductions

Introductions are normally related to the Bible book within which they appear, with testament introductions appearing at the start of a book group <div> but before the first book <div> and Bible introductions appearing before the first book group <div>.

However, in the past, testament and Bible introductions have been moved inside the first book of the testament or Bible. This was done because the SFM to OSIS conversion used to be unable to handle material which was not within a book.

Where a Bible or testament introduction does occur within book <div>, the following properties may be specified to cause the introduction to be placed correctly in the output.

“BibleIntro=True” :  
This indicates that the introduction to the first book of the Bible should be considered to be a whole Bible introduction.

“TestamentIntro=True” :  
This indicates that the introduction to the first book of each book group should be considered to be a testament introduction.

“CombinedIntros=True” :  
This enables the following behaviour. If a book introduction contains for the first book in a book group contains a title matching the book title as specified in convert.txt, then any introductory text following this is considered the book introduction, while any text preceding this is the Testament introduction.

Note that for OSIS files created with the latest version of osis-converters, Bible and testament introductions are no longer included within book introductions, so these the above properties should not be specified.

#### Book Subheadings

Occasionally a book title which appears in the OSIS file as a level 1 title is immediately followed by a level 2 title which is a sub-heading for the book. However, there are also instances where a level 2 title following a book title is not intended to be a book sub-heading. In order for such a level 2 title to be recognised, positioned and formatted a book sub-heading, convert.txt must contain the line “BookSubtitles=True”. Without this, the level 2 heading is considered to be a section heading.

#### Book Introductions

Books may have introductions which appear before chapter 1. By default, a title at the start of a book, following any book title and book subheading and before the start of the first chapter is the title for a book introduction. By default, any such heading will appear in the table of contents. To prevent such titles from being included in the table of contents, include the line “IntroInContents=False” in convert.txt. Specifying this setting also has the effect of preventing individual entries being generated in the table of contents for each <div> within a Bible or testament introduction.

#### Optional Line Breaks

By default, line breaks marked as optional in the OSIS will be ignored. To include such line breaks in the output, include the line “OptionalBreaks=True” in convert.txt

#### EPUB3

In order to make use of EPUB3 features supporting footnotes, convert.txt must contain the line “Epub3=True”. Where this line is not present, the traditional method of implementing footnotes using simple links between the footnote reference and the footnote text are used.

Not all readers support EPUB3, and in such readers footnote links may not work if “Epub3=True” is specified. On the other hand, for those readers which do support EPUB3, using the EPUB3 footnote support should give improved display of footnotes, with footnote text being displayed in a pop-up box or a separate area of the screen when the footnote reference is selected, rather than the main text display moving to the footnote text.

#### Glossary

The property GlossaryTitle may be used for the following purposes:

* To provide a title for a single glossary where no title is provided in the OSIS file – a title specified in the OSIS file will be used in preference to the title specified here.
* To provide a group title where there is more than one glossary. This is used only where Testament titles have been specified and there is more than one glossary. In this case, the title specified provides a title which will appear in the table of contents at the same level as the Testament titles, with the individual glossaries beneath it.

For example:

GlossaryTitle= Sözlük

The property GlossEntriesInToc controls whether individual glossary entries appear in the table of contents. By default, they do appear for EPUB and MOBI but not for FB2. Including the line “GlossEntriesInToc=False” results in individual entries not appearing in the table of contents for any output format. Including the line “GlossEntriesInToc=True” results in individual entries appearing in the table of contents for all output formats, including FB2. This requires the layout of glossary entries to be changed for FB2.

### 3.3 ebible.css

A copy of ebible.css is provided with the conversion tool. The ebible.css file used for a specific conversion will typically be an exact copy of this, but it is acceptable to use an edited version.

This file provides formatting information, and defines a number of classes.

It should be noted that h1 styling is used only for Testament titles, and h2 for book titles. All other level 1 titles use h3 styling and level 2 titles use h4 styling.

Among the classes defined are the following.

|  |  |
| --- | --- |
| **alt-var** | Applies to text marked as an alternate or variant reading. |
| **book-subtitle** | Applies to book subtitles, to distinguish them from section headings. |
| **canonical** | Applies to psalm canonical titles. |
| **first-para** | Used to avoid the normal indentation for the start of a paragraph immediately following a large chapter number. |
| **foreign** | Applies to text marked as foreign. |
| **dGlossary-entry** | Defines the spacing around a glossary entry. |
| **heading** | Applies to headings which are not titles, such a list heading |
| **psalm-div-heading** | Applies to psalm division titles, to distinguish them from normal section headings. |
| **transChange** | Applies to text marked as an addition or change to the translation of the original text. |
| **x-chapter-number** | Defines the style for large chapter numbers. |
| **x-chapter-title** | Applies to psalm and chapter titles, to distinguish them from normal section headings. |
| **x-glossary-link** | Defines a style used to highlight words which have corresponding glossary entries. |
| **x-introduction** | Applies to book introductory text, to distinguish it from the canonical text of the book. |
| **x-list-1** | Applies to list entries in introductions. |

The classes at the start of the file, from divineName to wordsOfJesus are provided for possible future use.

Most of the remaining classes are used to lay out poetical text.

Type attributes used in OSIS <p> elements must be defined as classes in ebible.css. An example of such a class in x-noindent, which suppresses the normal indenting of the first line of a paragraph.

Some of the styling, will not be effective for FB2. Styling which will not be effective includes:

* Setting background or text colour
* Styling which applies to spacing of elements
* Styling applied to section titles (i.e. testament, book and chapter titles), e.g. the styling in x‑chapter‑title.

## Running the Conversion Tool

The conversion tool is run by executing the osis2ebook.pl script. The format of the command line is:

osis2ebook.pl project-directory log-file inputs-directory OSIS-file output-format [<input-type> [<cover-image>]]

The parameters are as follows.

* **Project-directory.** The path to the osis-converters project directory.
* **Log-file.** The path to the osis-converters log file.
* **Directory.** The path of the directory containing OSIS file, convert.txt.
* **OSIS-file.** The name of the input OSIS file, including the .xml extension.
* **Output-format.** The output format to be written – “epub”, “mobi” or “fb2”. In theory, any output format supported by Calibre can be specified, but other formats may not produce satisfactory results.
* **Input-type.** This optional parameter is provided for future use. At present, if specified, it must be “B” or “Bible”.
* **Cover-image.** This optional parameter is the path of a file containing a cover image. If a relative path is specified, this should be relative to the directory specified in the first parameter. If no cover image is provided and the output format is epub, a default cover image will be generated.

The output file is created in the same directory as the OSIS file, and has the same name, but with the .xml extension replaced with the appropriate extension for the output format.