Standard for Literacy Content in Paratext 0.5

This document presents a standard for the PBT-PNG branch for storing literacy materials such as shell books and original stories in Paratext. The motivation for developing these standards is that many have begun to store shell books and original stories in the Extra books of Paratext projects (XXA-XXG), but there has been no consistency in the formatting, which can lead to much confusion and inefficiency. To reduce the potential for confusion and to increase efficiency, this standard has been proposed.

# Goals for Standard

These standards have been created to meet the following needs:

1. There is a need to have a single, secure repository of literacy materials for each language. A single literacy work, e.g. a shell book, is frequently published in many different formats and sizes, all with the same textual content. In addition to these differing formats, there may be multiple revisions of the work with different spellings. There needs to be a single, authoritative copy of the latest version of the text of each work in an agreed upon location, that is easily accessible by many. Having shell books in a Paratext project can meet this need.
2. There is a need to spellcheck shell books, original stories, and other literacy materials according to the same spelling standard as Scripture. Paratext provides this capability.
3. There is a need for literacy workers who work with multiple languages to apply the same practices and tools to all the languages that they work with.
4. The standard needs to accommodate the unique features of literacy materials, but allow literacy workers to benefit from the many tools built into Paratext. For example, the standards must be compatible with the Word List tool and Paratext’s printing functions.
5. Since shell books and other literacy materials will be published with a tool other than Paratext—frequently Publisher—it is not necessary that the materials have a polished appearance in Paratext. It is only necessary that the components can be easily recognized and can be easily ported to publishing software.

# Definitions and Conventions

**book** - a literacy work, typically a shell book or an original story published as a booklet. It could potentially also be a song book or other literacy content. To prevent confusion, any usage of the work book that does not correspond to this definition will be accompanied by a descriptive word to make its meaning clear.

**marker** or SFM marker – a back slash character (“\”) plus one to three alphanumeric characters that immediately follow it. Examples are: \rem, \c, and \m.

**category –** a predefined category for shell books, such as Animal Stories or Health. Other groups of books also have categories.

**category code –** a predefined abbreviation of a category. Usually the category code consists of two capital letters, but there are a few exceptions.[[1]](#endnote-1) Examples of category codes are “AG” for agriculture and “AS” for animal stories. Category codes may be assigned to other groups of books other than shell books. For example, “OR” is the category code assigned to original stories.

**book number –** a book number is a sequence of letters and numerals that uniquely identify a shell book or other literacy work. Book numbers are assigned by the publications and literacy team. A typical book number consists of a two letter shell book category code followed by a four digit number with no intervening space or punctuation, for example AS0001 and TG0017. There are some exceptions and extensions to this norm. Many of the exceptions and extensions used by the literacy and publications team are described in Appendix A.

# Standard Defined

Following is the standard for literacy works in Paratext. It has been specifically designed to meet the needs of shell books, but can be used for other kinds of books. This standard applies only to the text of a book that is in Paratext. It does not apply to illustrations, publication information, and other components of a published book that is stored outside of Paratext.

**XXA –** The Paratext book named *Extra A* (XXA) will be used to store books in a Paratext project. Literacy materials may be stored in other Paratext books, but this standard only applies to books in XXA.

**Basic organization –** The basic organization will be by category, with one chapter being assigned to each category of shell book, and one verse number to each book in the category. Currently there are 31 shell book categories being used by the PBT-PNG literacy and publications team, so chapters 1-31 will be allocated to these categories. Other groups of books that have a common category code will be assigned to higher chapter numbers, with one category code per chapter. Currently chapter 32 has been allocated to Original Stories (category code OR) and chapter 33 to Storybooks (category code ST).

**Non-vernacular text –** One of the primary purposes of putting books into Paratext is spell checking, so it is important to keep non-vernacular text—category codes, book numbers, English titles, authors, etc.—out of the Paratext Word List. Since the Word List tool ignores text after \rem markers, all non-vernacular text and metadata will be placed after \rem markers.

**Markers before books in a chapter –** After the chapter number and before the first book (verse) will be one line beginning with \rem that contains the category code followed by the category in parentheses[[2]](#endnote-2). Examples:

\rem AG (Agriculture)

\rem BG (Bible in PNG)

**Book header –** There will be one book per verse. Each shell book or original story will begin with one verse marker, \v followed by verse number. (The verse number has no significance to the book, other than indicating its boundaries and position in the chapter.) Since Paratext requires that a verse marker occur after some kind of paragraph marker (\p, \m, \q1, \q2, etc.), the first line in a book will be \m \v [*number*]. The verse number may optionally have a single letter verse segment indicator[[3]](#endnote-3) as a suffix (\v 3b). The segment markers are ideal for marking multiple versions of the same book.

The second line in the book will begin with the marker \rem, followed by the book number. Immediately following the book number will be the title of the book in English.

Following are examples of book headers (These examples include exceptions and extensions of book numbers described in Appendix A.):

\m \v 1

\rem AG0001 The Seed

\m \v 1a

\rem AG0001sla The Seed (simplified)

\m \v 14

\rem AS0015 Rat and Snake

\m \v 11

\rem OR-AG0011 The Dog and Possum

**Metadata –** After the book header may be any number of \rem markers, each indicating some kind of metadata. There may be labeled metadata and freeform metadata. Labeled metadata will consist of a \rem marker followed by a label followed by the metadata content. The label typically will consist of a single word and will be terminated by a colon (“:”). Following are examples of labeled metadata:

\rem author: Willy Kanoba

\rem ISBN: 972-97421-66-2

Following are some suggested labels and their uses:

author: original author of the book

artist: original artist of the book

adaptor: person who adapted a book authored by someone else by translating it into another language and/or modifying the content.

status: spell checking status of the book

ISBN: ISBN number

dialect:

curriculum:

publication history:

Some special labels:

developer: This label is designed to introduce publication information provided by the original developer. It may contain copyright information, author, artist, dates, venue where developed, ISBN, etc. It is designed to contain the publication information in the Developer column of the shell book catalog.

permission date: Following this label is the date that a permission form from a contributor was filled out and filed in the Publications office. This label occurs immediately after the name of a contributor in an author, artist, or adaptor field. Following are examples:

\rem author: Willy Kanoba, permission date: 11/01/2006

\rem artist: Sera Witaki, permission date: 13/06/2015

\rem adaptor: Kathi Olathi, permission date: 25/10/1999

Freeform metadata will begin with a \rem marker, and have no other specifications. An example of freeform metadata is as follows.

\rem Translated by Sochi Manut in 2007 writers workshop.

**Vernacular title –** The \s marker is used to mark the vernacular title of the book. In Paratext \s is a section header and formatted as centered and in bold.

**Pages –** Shell books and original stories are rendered on multiple pages in Publisher. The page breaks need to be marked in Paratext. There is no ideal marker for this[[4]](#endnote-4), so “\m ” followed by a page number will be used to mark the beginning of a page. The page number will consist of “p.” followed by a number. There will be one space and no punctuation following the page number. Here an example:

\s [vernacular title]

\m p.3 [vernacular text, 1st page]

\m p.5. [vernacular text, 2nd page]

\m p.7 …

**Paragraphs –** Many books have no more than one paragraph per page, so for most books, marking paragraphs is irrelevant. Some books do have multiple paragraphs on a single page, so for these books it is appropriate to mark the paragraph breaks. The marker \p is the normal paragraph marker in Paratext. So for a book with multiple paragraphs on a page, the formatting would look something like this:

\s [*vernacular title*]

\m p.3 [*vernacular text, 1st page*]

\p [*vernacular text, 2nd paragraph*]

\p [*vernacular text, 3rd paragraph*]

\m p.5 [*vernacular text, 2nd page*]

\m p.7 …

**Speech Balloons –** Some shell books have speech balloon with text in a picture, usually on an even numbered page. A speech balloon will be preceded by the marker \pm. Here is an example:

\s [*vernacular title*]

\m p.3 [*vernacular text, 1st page*]

\pm p.4 [*vernacular text in speech balloon*]

\m p.5 [*vernacular text, 2nd page*]

\m p.7 …

**Common Repository for shells -** A copy of all the shell book shells will be stored in the Extra book XXA of a common Paratext project that all users can have access to. Since PBTPShel already contains Scripture and front matter shells used by the branch, it will also be used to store shell book shells. These shells will follow the standard described in this document.

When literacy work starts for a new language project, it is suggested that the Paratext project administrator import the entire XXA book from PBTPShel into the language project.

# Appendix A – Book number conventions

The book number conventions for shell books used by the literacy and publications team are well established and are not subject to change. These conventions include exceptions and extensions to the typical book number structure of two capital letters followed by four digits. Since changing them is not an option, it is prudent that Standard for Literacy Content in Paratext be able to accommodate these conventions.

Following is a description of some of these conventions. These conventions are not part of the proposed standard, but included as a courtesy to the reader.

* The category code for Scripture is a single letter “S”. Example: “S0019”
* The category code for Kartidaya is “S-“ and is followed by a two digit number. Example: “S‑04”
* Some shell books have two sets of pictures for the same story. The pictures in one set are from D.C.Cook and the pictures in the other set are from another publisher. The book numbers for shell books from D.C.Cook are preceded by a lower case “c”, while shell books with pictures from another publisher are not preceded by lower case “c”. Examples: “OT0003”, “cOT0003”, “NT0027”, and “cNT0026”.
* The book number for stories in category OR (Original Stories) is composed of “OR-” followed by a two letter code for the language, followed by a four digit number. Examples: “OR-MB0003” (Mborena Kam) and “OR-AG0022” (Akukem)
* Some shell book numbers have suffixes that indicate a variety of things. Here are some examples:
  + CO0001D – A single shell book CO0001 was divided into five parts indicated by suffixes A – E.
  + FG0026sl – A simplified version of FG0026
  + NG0007px2
  + OT0001slb – OT0001 has two simplified versions, OT0001sla, and OT0001slb.

# Appendix B – Helps for converting non-standard page numbers to standard page numbers.

There are a lot of shell book texts already in Paratext, and they exhibit a wide variety of page numbering formats. To convert all these to the standard page numbering system could potentially involve many hours of tedious work. However, in many cases this work can be sped up by using Paratext’s Find/Replace tool, taking advantage of its built-in *regular expression* functionality.

Below are several scenarios, each featuring one kind of formatting that does not follow the page number standard along with instructions for converting the non-standard page number formatting to a format consistent with the standard. Each scenario contains a table with three cells: 1) a sample text with non-standard formatting before doing Find/Replace, 2) what the text would look like after the Find/Replace is applied, and 3) helps for performing the change.

The helps may be in one of two forms. The first form consists simply a list of other scenarios to perform in order to achieve the result. The second form consists of Find and Replace texts for use with the Paratext Find/Replace tool.

Here are guidelines for using the Find/Replace tool:

* Open the tool using **Edit > Replace** or **CTRL-H.**
* Copy the text between the quotation marks in the **Find** column to the **Find What** box in the Find/Replace tool. Include *all* the text between the quotation marks, because here may be space characters at the beginning or end of the Find text, and you must copy these also.
* Copy the text between the quotation marks in the **Replace** column to the **Replace With** box in the Find/Replace tool. Include *all* the text between the quotation marks, because there may be space characters at the beginning or end of the Replace text.
* Limit the search to one Paratext book to prevent undesirable changes from occurring in Paratext Scripture books.
* Consider limiting changes to the current chapter by ticking the Current Chapter box to prevent unwanted changes from occurring in other chapters.
* As usual, after you click the Replace button, Paratext will give you an opportunity to accept or reject each change.

## Scenario #1: Inserting “p.” at the beginning of page numbers followed by a period

|  |  |  |
| --- | --- | --- |
| Before | After | Steps |
| \m 3. Liko pani dor  \m 5. Liko pani dor  \m 7. Liko pani dor | \m p.3 Liko pani dor  \m p.5 Liko pani dor  \m p.7 Liko pani dor | Do Scenario #2  Do Scenario #3 |

## Scenario #2: Removing periods after simple page numbers

|  |  |  |  |
| --- | --- | --- | --- |
| Before | After | Find | Replace |
| \m 3. Liko pani dor  \m 5. Liko pani dor  \m 7. Liko pani dor | \m 3 Liko pani dor  \m 5 Liko pani dor  \m 7 Liko pani dor | **“regex:(?<=\\m \d+)\. ”** | “ “ |

## Scenario #3: Inserting “p.” at the beginning of simple page numbers

|  |  |  |  |
| --- | --- | --- | --- |
| Before | After | Find | Replace |
| \m 3 Liko pani dor  \m 5 Liko pani dor  \m 7 Liko pani dor | \m p.3 Liko pani dor  \m p.5 Liko pani dor  \m p.7 Liko pani dor | “**regex:\\m (?=\d+ )**” | “**\m p.**” |

## Scenario #4: Converting verse numbers to page numbers

|  |  |  |  |
| --- | --- | --- | --- |
| Before | After | Find | Replace |
| \v 3 Liko pani dor  \v 5 Liko pani dor  \v 7 Liko pani dor | \m p.3 Liko pani dor  \m p.5 Liko pani dor  \m p.7 Liko pani dor | “**\v** ” | “**\m p.**” |

## Scenario #5: Converting simple page numbers in running text to standard page numbers

|  |  |  |  |
| --- | --- | --- | --- |
| Before | After | Find | Replace |
| 3. Buli tav nor. 5. Sisi dila po kiki? 7. “Loso nuar tina.” 9. Ui na san ti do. | \m p.3 Buli tav nor.  \m p.5 Sisi dila po kiki?  \m p.7 “Loso nuar tina.”  \m p.9 Ui na san ti do. | “**regex:(?<=[\.\!\?][”\"]?)\s+(?=\d+\. )**” | “ **\m p.**” |

1. See Appendix A. [↑](#endnote-ref-1)
2. The purpose of the parentheses is to give the \rem line a unique formatting that can be readily recognized by software. [↑](#endnote-ref-2)
3. The characters that can be used as verse segment indicators is set in the language settings of each project (Project > Language Settings > Other Characters tab > Verse segment characters). In most PBT-PNG projects, the available characters are ‘a’ through ‘z’. [↑](#endnote-ref-3)
4. Paratext does have a page breaker marker, \pb, but when the Print Draft function of Paratext encounters a \pb it immediately advances to the next physical page. If \pb were to be used to mark pages in books, a printout of a single book would be spread out over several sheets of paper. [↑](#endnote-ref-4)