GLOBAL BOOK ALLIANCE

Contributing Materials to the Global Digital Library –

Conversion from PDF to MS Word

Reading materials are usually produced using software such as Adobe InDesign, Microsoft Word or Publisher, or Bloom.

These programs allow for the content to be exported into various formats, including ePUB which is used on the Global Digital Library as well as PDF. PDF is a preferred format for conserving the look and feel of a page, is used globally and accessible on most devices.

PDF documents are flattened versions of the originals. However in conserving the format of the page, a lot of digital information is lost, which has implications for usability and accessibility. Such files cannot be edited without a special program, such as Adobe Acrobat.

Since materials are not created directly in PDF format, it means there is an original version in InDesign, Word etc. It is recommended that the original files used to create the storybooks are accessed since this will make the process of conversion into ePUB simpler and less time-consuming. The original files will have selectable, copyable text, and will have options for page order, text editing and high-resolution images.

However, in some cases, the original files are not accessible.

In these cases, the materials need to be re-created in a book editing platform: Bloom, InDesign, and MS Word are just a few of the many options available. And from there, the document must be converted into its final format, the ePUB file.

Provided here are instructions for how to:
1. Recreate PDF storybooks in MS Word
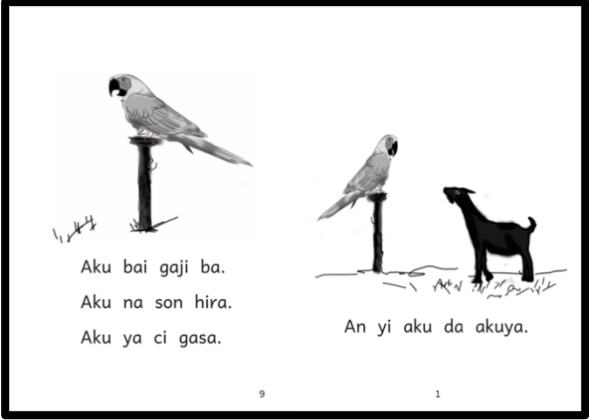
RE-CREATE THE STORYBOOKS IN MS WORD

To re-create the files in MS Word:

1. Open the PDF of the book.

2. Open a blank Microsoft Word document.

3. For each page, use your computer's screenshot application to capture each image and paste it into the MS Word Document.

4. Manually type out the words that belong on that page, using the MS text tool.

    Ensure:
    a. Pages appear in the order in which they are meant to be read, and just one page of the storybook is shown on each page in MS Word.
    b. Page numbers are added (View – Header and Footer – Insert – page numbers – alignment - center)
    c. The title of the books is in "Title" formatted text (Styles – Title)
    d. The main text for the book is in "Normal" text style (Styles – Normal)
    e. Formatting matches the original book – for early grade readers, the image should take up the bulk of the page, with large text (28-35 pt) below the image.
    f. Images are high resolution enough to be comfortably read on an e-reader – at least 200 x 200 pixels.
    g. Each story has a one-inch margin (Format – Document)
    h. The font used is accessible. Examples include: Arial, Helvetica, Lucida Sans, Tahoma, and Verdana (recommend 28-35 pt font for story pages, 12 pt font for other text — for example, about the development of the books, rights and licensing).

5. When you have re-created the book into MS Word, you are ready to convert the story into an ePUB file!

6. For help and guidance, reach out to comms@globalbookalliance.org

Example - Page formatting

Before Re-creation	After Recreation
Before Re-creation	**After Recreation**
In this example, the book was created for print purposes, and thus page 1 and 9 are printed on the same page. This is because the book is meant to be printed and bound with other pages stacked on top of it. Digitally, however, the book will not be readable in this format.	After re-creating the book into the Word format, pages have been separated into single spreads and can be viewed in order. Shown here are pages 1 and 2 of the book.

Example – When words are part of images

Before Re-creation	After Recreation
Before Re-creation	**After Recreation**
In this example, the words are part of the picture. If possible, remove the words from the photo using photo editing software, and re-create the text in an MS Word text box.	A screenshot was taken of the image without the text, and then text was added and rotated using MS Word's Text Box feature (Insert – Text Box)