National Library of Scotland Wikisource Project: Proofreading Guide

Getting started on Wikisource

1. In order to correct our digital collections' OCR on the Wikisource platform you will require a Wikipedia account.

2. If you **don't** already have a Wikipedia / Wikisource account, take the following steps:
   a. Go to the Wikisource website https://en.wikisource.org/wiki/Main_Page and create your account by clicking "create account" at the top right.
   b. Once you have created an account, your username will be in red at the top of the Wikisource page. Click on this and then press the blue "create your own" button to set up your own user page.
   c. Next, write some text introducing yourself.
   d. Next, we need to change a setting in our account which will make the proofreading process easier later. This is to activate Google OCR in our Wikisource editor.
      i. To do this, click on "preferences" at the top right of the screen when you are logged into Wikisource, click on the "gadgets" tab then scroll down and check the box in the "Developments (Beta)" area called "Google OCR". Then click on save.
   ii. (Then when you go in to proofread an item, there should be a rainbow coloured OCR button in the text editor screen. Click this button and it will replace the automatically generated OCR with a higher quality Google-generated OCR.)

Getting ready for proofreading

3. Before you can start editing the item showing on the Wikisource index page, you'll need to log into your Wikisource account at the top right of the screen using your Wikipedia / Wikisource credentials.

4. Each book's index page should look something like the screenshot below. As you can see, a list of the item's pages is shown underneath the thumbnail image. All of these pages are written in red, meaning that none of them have been proofread by a user yet.
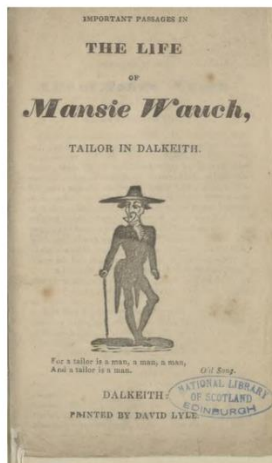
5. In order to start proofreading a page's OCR, click on one of the red pages listed underneath the item's thumbnail image. (It is recommended that you start with page 1 and work your way through the item.)

6. Your screen should now show the following, with the original image on the right, and the pre-populated OCR transcription on the left:

Proofreading guide – mandatory steps

7. Go through the OCR transcription, checking for any cases where this does not match the original text printed on the image to the right. You can correct any mismatches by directly editing the OCR within the text field on the left.

8. **Spelling and punctuation:** Retain the original item's spelling and punctuation. Do not modernise or correct any spelling other than the long S (see step 13 below), even if the original text includes typos.

   Where such typos are auto corrected in the OCR, please change this back to the way they are printed in the original item. Particular letters to look out for are e/c/o and h/b as these are commonly mixed up.

   Where words or sentences are printed with unexpected spaces, please retain those.

9. **Long S (ſ):** <u>REVISED 11th MAY</u>: Now that we are using Google OCR (as shown in Step 2d above). This OCR is much better at identifying the long-s letter (e.g. the word "season" can appear as "ſeaſon" in the text) so when you start your next new book, please **include** rather than remove the long-s letter when proofreading. The new OCR will generally detect these well, but occasionally it will read the letter as "s" or an "f". When this happens, just copy the "ſ" from somewhere else in the text to replace it.

10. **Word split over two pages:** Where words are split over two pages, write the full word (e.g. "generations") on *either* the first or second page – instead of retaining the split (e.g. "gene-" on the first page and "-rations" on the second).
    a. With **Words split over two lines**, connect these back into one word, unless the word is hyphenated in the original text.

11. **Line breaks:** Rules for line breaks differ between prose and verse texts, as well as on title pages. See guidelines for each below:

    A) Prose: On pages with prose, you will see that the pre-populated OCR transcription will look something like the image below, where there is a new line for every new line in the original print:
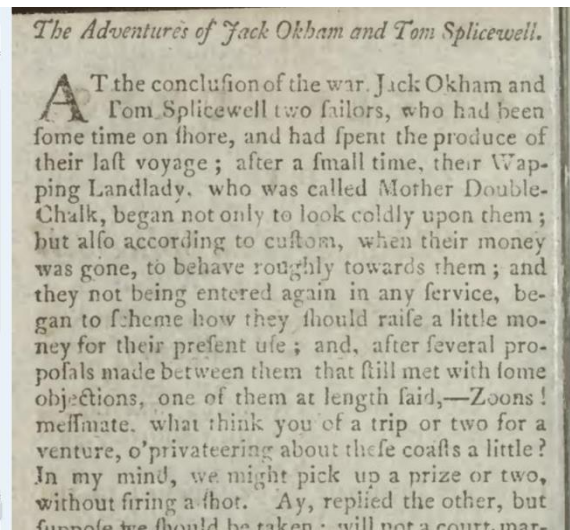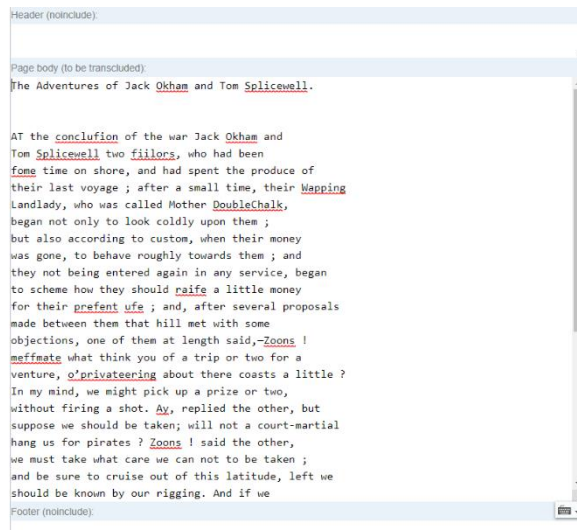
We will need to remove these line breaks (even though they will look fine in the Preview/Published page as mentioned above, they will not look right in the exported txt file). Instead, the transcription should look more like this:
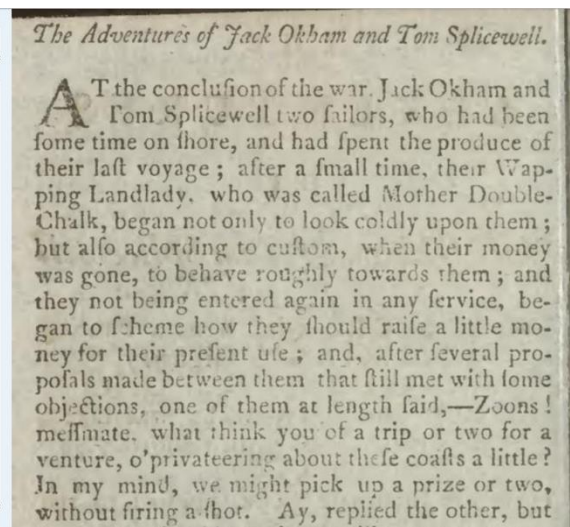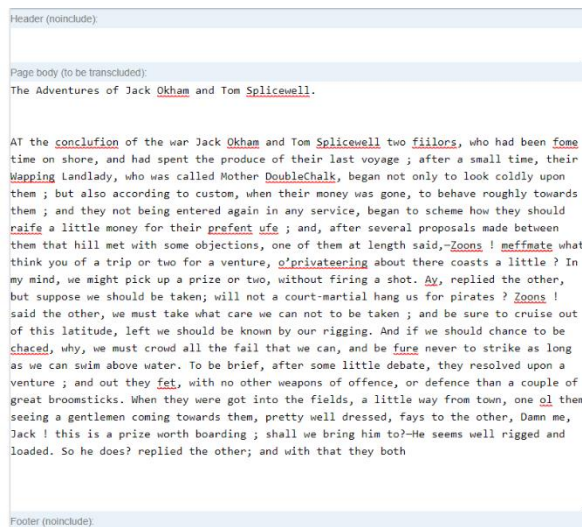
To start a new paragraph, press the ENTER key twice.

B) Verse/Poetry:

    a. Remove line breaks on overflow lines. By this we mean turning:

<div align="center">

Shall I compare thee to a summer's

day

</div>

    Into:

<div align="center">

Shall I compare thee to a summer's day

</div>

    b. Type `<br />` at the end of each line in a stanza (a group of lines), except for the last line.

    c. To start a new stanza, press the ENTER key twice.

    d. If a line in the original poem is indented, type `{{em}}` at the beginning of the line

This is an example of a common layout for poems:

How could you say my face was fair,
    And yet that face forsake?
How could you win my virgin heart,
    Yet leave that heart to break?
How could you promise love to me,
    And not that promise keep?
Why did you swear mine eyes were bright,
    Yet leave those eyes to weep?

How could you say my lip was sweet,
    And made the scarlet pale?
And why did I, young witless maid,
    Believe the flatt'ring tale?
That face, alas! no more is fair;
    That lip no longer red;
Dark are mine eyes, now clos'd in death,
    And ev'ry charm is fled.

This text should look like this in the transclusion and preview panes:

```
How could you say my face was fair,<br />
{{em}}And yet that face forsake?<br />
How could you win my virgin heart,<br />
{{em}}Yet leave that heart to break?<br />
How could you promise love to me,<br />
{{em}}And not that promise keep?<br />
Why did you swear mine eyes were bright,<br />
{{em}}Yet leave those eyes to weep?

How could you say my lip was sweet,<br />
{{em}}And made the scarlet pale?<br />
And why did I, young witless maid,<br />
{{em}}Believe the flatt'ring  tale?<br />
That face, alas ! no more is fair;<br />
{{em}}That lip no longer red;<br />
Dark are mine eyes, now clos'd in death,<br />
{{em}}And ev'ry charm is fled.
```

How could you say my face was fair,
    And yet that face forsake?
How could you win my virgin heart,
    Yet leave that heart to break?
How could you promise love to me,
    And not that promise keep?
Why did you swear mine eyes were bright,
    Yet leave those eyes to weep?

How could you say my lip was sweet,
    And made the scarlet pale?
And why did I, young witless maid,
    Believe the flatt'ring tale?
That face, alas ! no more is fair;
    That lip no longer red;
Dark are mine eyes, now clos'd in death,
    And ev'ry charm is fled.

C) Title pages: You may find that for some several lines in the OCR transcription. Corre

12. **Illegible text:** Where text is illegible (e.g. through being faint or cut off at the page edges) you should add missing letters or words to the transcript if you can confidently tell what they should be.

If you cannot do this confidently, choose *one* of the following options:

   a. Write {{illegible}} as a placeholder.
       i. Note: This can be used to represent single letters, whole words, or even phrases. You do *not* need to insert this placeholder multiple times to express that a row of consecutive words is missing / illegible.

b. Leave those missing letters / words out. This might be your preferred choice if you are not comfortable using the tags / mark-up. In this case, a colleague will insert them at the later 'validation' stage.

13. **Very blurry images:** You may find that some images or parts of images are unusually blurry. This is most likely due to an error within the file rather than due to the original item being blurry

14. **Blank pages:** You may come across pages which are blank but show text bleeding through from the other page side. This often results in the pre-populated OCR displaying random letters and symbols. Please delete all of these and follow the guidance outlined under step 29 on how to tag this page before saving it.

15. **Headers**: On many items, you may see headers and page numbers on each page. Often the pre-populated OCR will place this text in the main box:
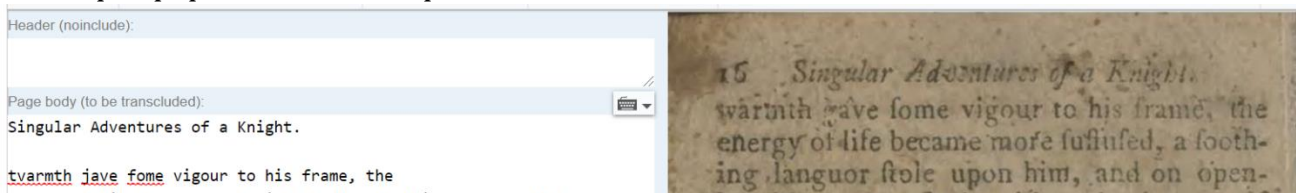
This needs to be moved into the Header (noinclude) box above it instead, otherwise when the text from one page is joined up with that of the next, the numbers and headers will be included in amongst the main content of the page.

Proofreading guide – optional steps

The following steps are very helpful to improve user experience on the Wikimedia platform. However, they are only optional at this proofreading stage. You are welcome to apply them while working on your pages and items, or you can choose to ignore them in which case they will be picked up by a colleague at the later 'validation' stage.

16. **Alignment:** When you encounter text which is centred or aligned to the right, as far as possible, you should match the OCR to the image.

    a. To centre text: {{center| Text goes here}}
       To align right: <p align="right">Text goes here</p>

17. **Basic formatting:** When text appears which is bold, italicised, or of differing sizes in the original, you can use the editing bar above the transcription pane (you may need to click Advanced to see the second row of the toolbar):
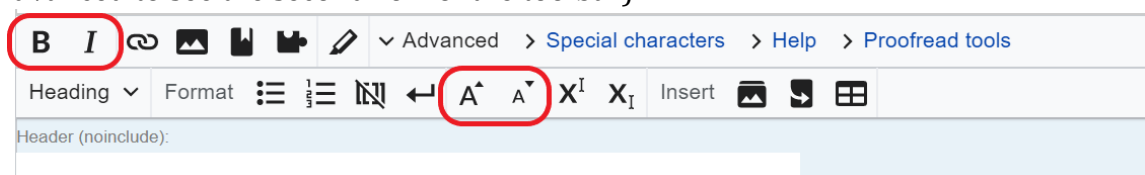
The options to make text **Bold** or *italic*, or change the text size are highlighted in the screenshot of the editing bar above. In order to apply any formatting to your transcription:

a. Select the text you want to format
b. Click the relevant button on the toolbar
c. The transcription should now have the tag applied before and after the text you selected:

The example above shows what adding the *italic* tag looks like, these mark ups will appear in the editing page, but will translate to formatted text in the preview.

18. **Special Characters:** Should you be working on a page with a more complex layout or unusual characters, you may want to have a look at formatting options available via the drop-down menu within the editing bar, as shown in the screenshot below.
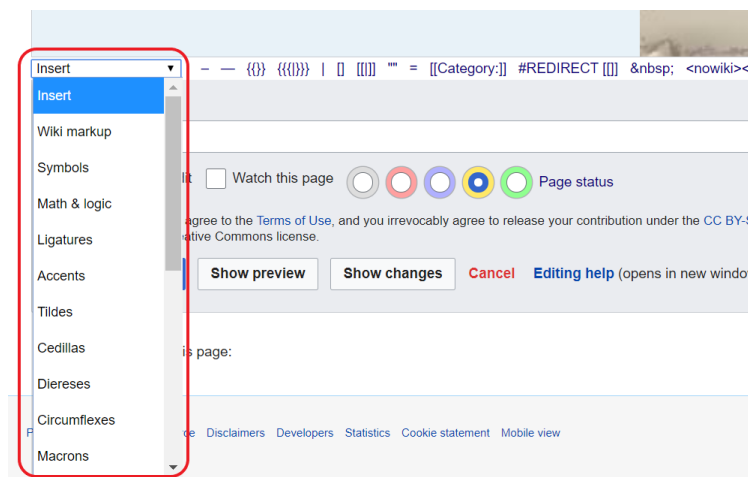
Note: the editing bar, including this drop-down menu, might be displayed below your editing box and image, instead of above.

19. **Columns:** If you encounter a text with tables, there is a useful help page at https://en.wikisource.org/wiki/Template:Columns.

For example, this template will create a simple table, the result of which is seen below:

{{columns|col1 = Simplest
instantiation
of|col2 = this
template}}

Simplest instantiation of	this template

You can add as many columns as you need by adding |col3 =, |col4 =, etc, and more rows by adding line breaks.

20. **Pre-populated mark-up:** You may find that the page you're proofreading includes mark-up symbols within the pre-populated OCR. This is mainly the case on title pages. Please retain this mark-up (unless you are confident that the mark-up is not correct) and only correct the text within. If you are not comfortable with pre-populated mark-up symbols you can save the page tagged with the lilac ("problematic") button (see step 29 below) and move to the next page.
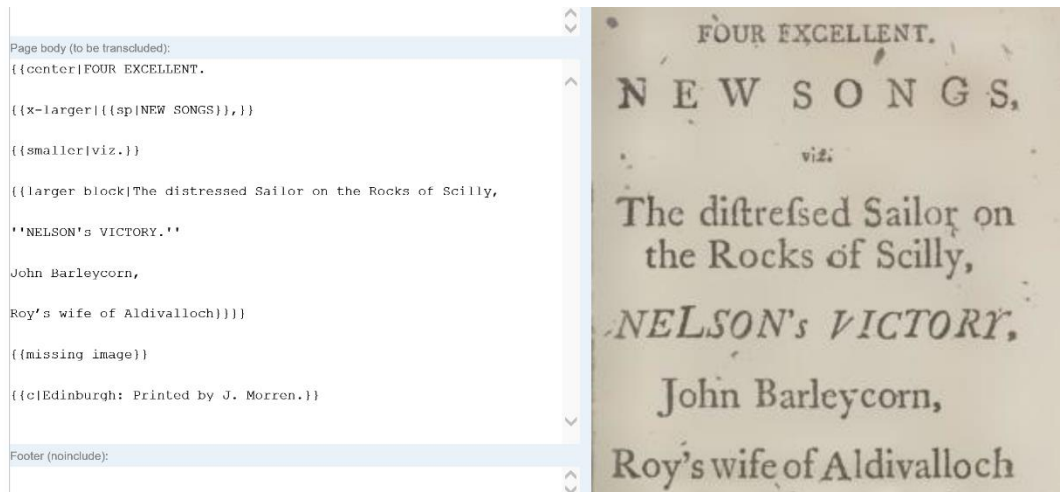
21. **Image/picture on original page:** Sometimes you may come across a picture or drawing. In order to reflect this in the transcription, we need to add a `{{missing image}}` tag to the transcription, making sure to leave blank lines above and below it. This should look like this:
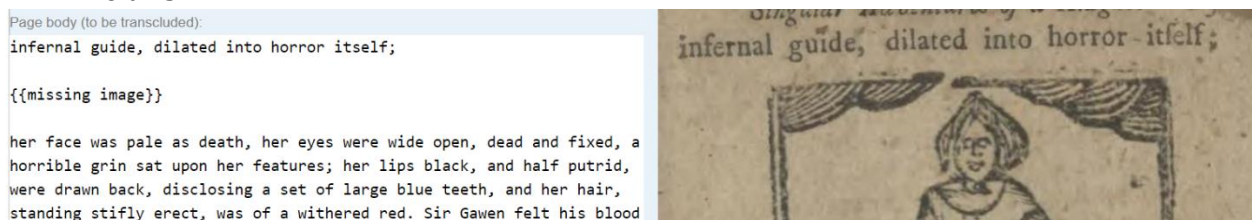
And on the preview/published page, it should look like this:


Saving your edits

22. Once you have proofread a page and feel you have done all you can or wish to pause and return to the page later, scroll to the bottom of the page and click the white "Show Preview" button to make sure it looks as you expect.
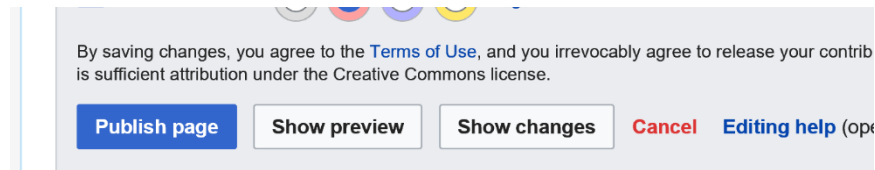
23. This will take you to a screen which looks like the screengrab below. This screen allows you to carry out further edits in the text box to the left of the image if need be.
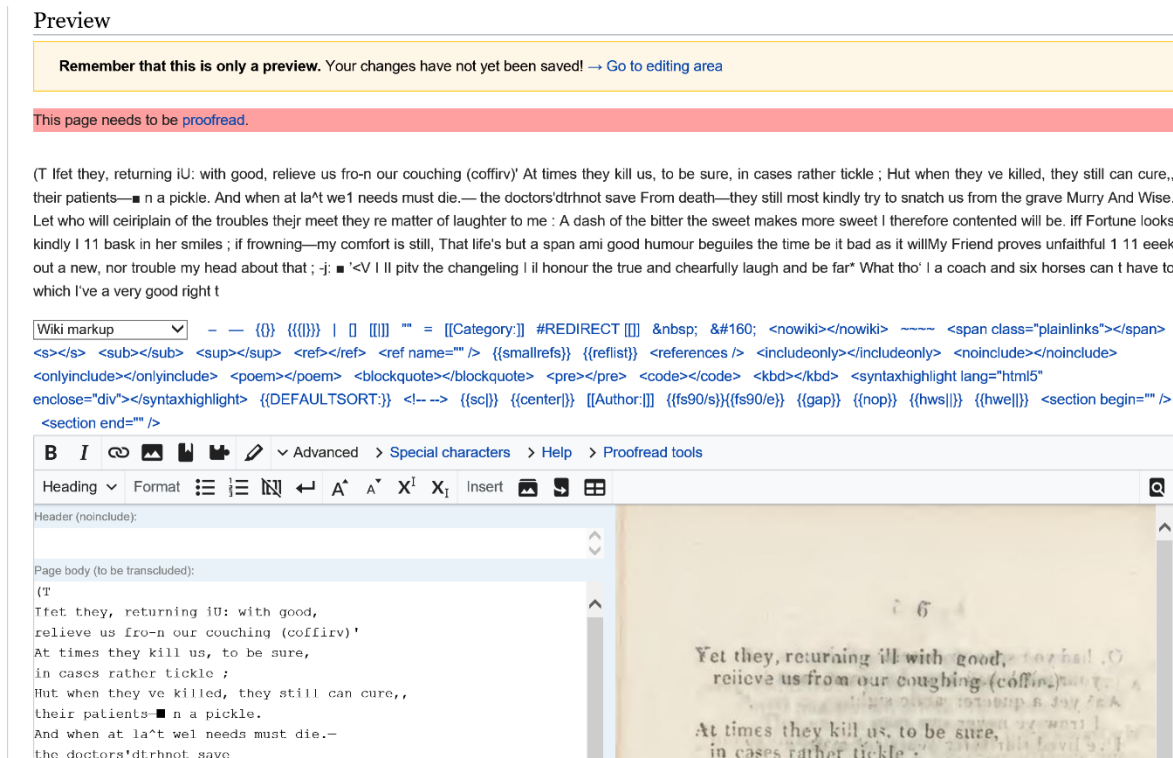
24. Once you are happy that the preview looks as you expect, you are ready to save your page, which is done in a **two-step** process as explained below.

25. As the **first step**, you need to record your page's status by tagging it with one of the four coloured circles that are displayed at the bottom of your screen.
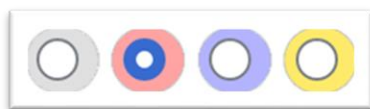

The circles are used in the following way:

a. Use the **GREY** ("without text") circle to indicate that the page is blank.
b. Use the **RED** ("not proofread") circle to indicate that you are not yet finished proofreading this page and want to return to it at a later stage.
c. Use the **LILAC** ("problematic") circle to indicate that the page (or parts of it) were too tricky to work on.

d. Use the **YELLOW** ("proofread") circle to indicate that you are finished proofreading this page and the page is neither blank nor tricky.
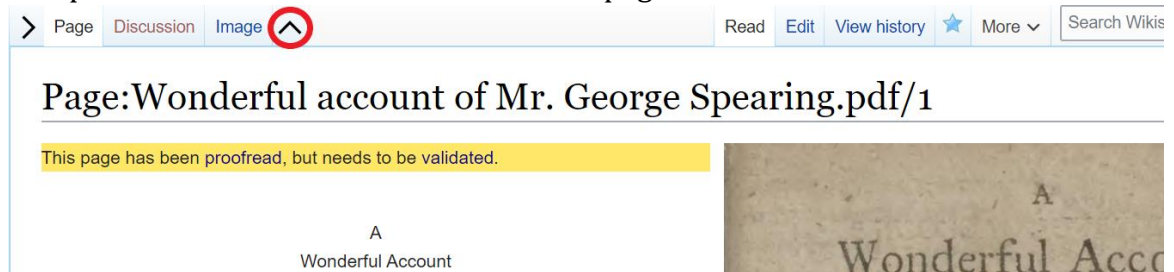
26. As the **second step**, press the blue "publish page" button to save your page.
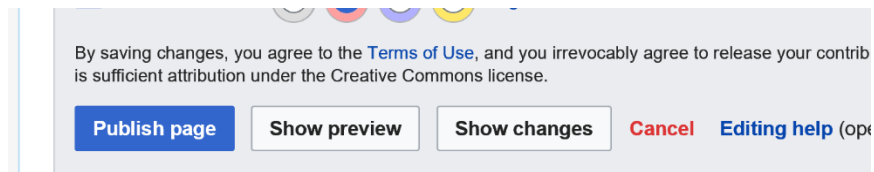

Moving to the next page

27. You can move to the next (or any other) page for proofreading in two ways:

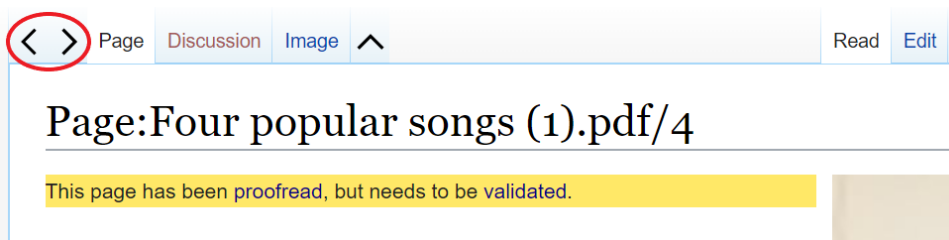a. Option 1: Click on the ^ arrow above the page title to return to the item's index.

From there you can go to a new page by selecting it from the list of pages under the image.

You will notice that the pages you've previously worked on are now highlighted in the colours you tagged them as (in the case of the below screenshot they are all yellow).
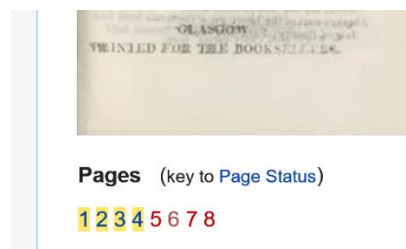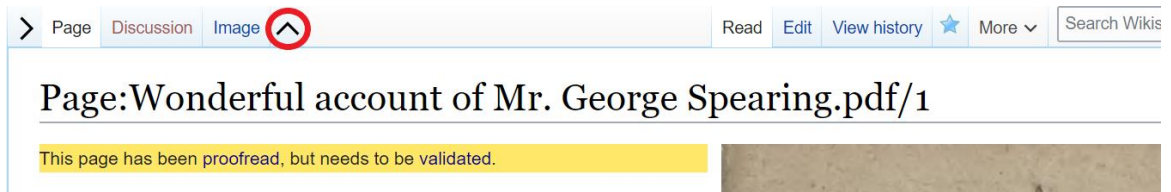

b. Option 2: Click on the < or > arrow above the page title to navigate to the previous or next page.

Re-editing a previous page

If you wish to return to a page which you have proofread before, you need to go to the item's index page. You can reach that page if you currently are on a different page within the item then click on the ^ arrow above the page (see screengrab below)

28. To re-start proofreading your page, click the "Edit" tab on the top right of the page to pick up where you left off.
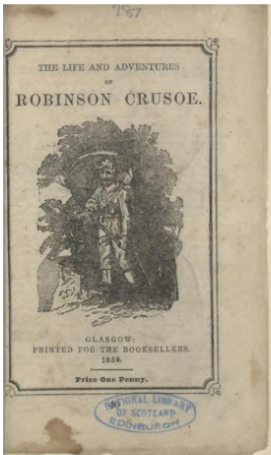

Finishing an item and starting a new one

29. On Wikisource, if all your pages are now yellow, set the Index as 'To be validated'. To do this, go to your Index page (shown below), then to the 'Edit' function in the top right corner.

A list of options will appear, so scroll down until you see the 'Progress' box which will currently say 'To be proofread'. Using the dropdown list, (shown below) select 'To be validated' then click on Publish changes at the bottom of the page.

30. You are now ready to move on to the next item by following this guide's instructions from step 6 onwards.