NYINGARN

User Guide

December 2024

Contents

1. Logging in to Nyingarn

The Nyingarn Workspace is accessed at https://workspace.nyingarn.net/. There are three options you can use to enter the Nyingarn Workspace. Before you can login, a Nyingarn administrator will need to add your email address to the Workspace as a registered user.
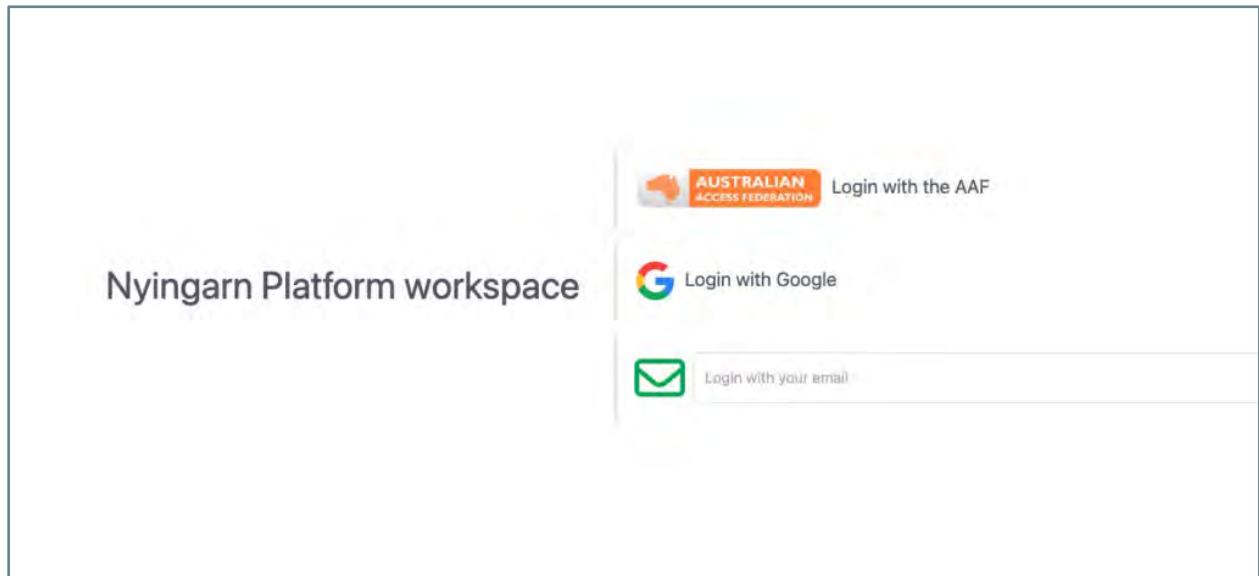

1. Through the Australian Access Federation (AAF) with an Australian university email address.

2. Through a Google Gmail account.

3. If you don't have a Gmail or AAF email address, you can use any other email address by entering your email into the box after the green envelope. After entering your address, a login code will be sent to your email. Depending on your email service it could take a few minutes to arrive and it might be in your spam folder. You can close this window, as the link in your email will open in a new window.

2. Uploading a manuscript

2.1 Creating an item

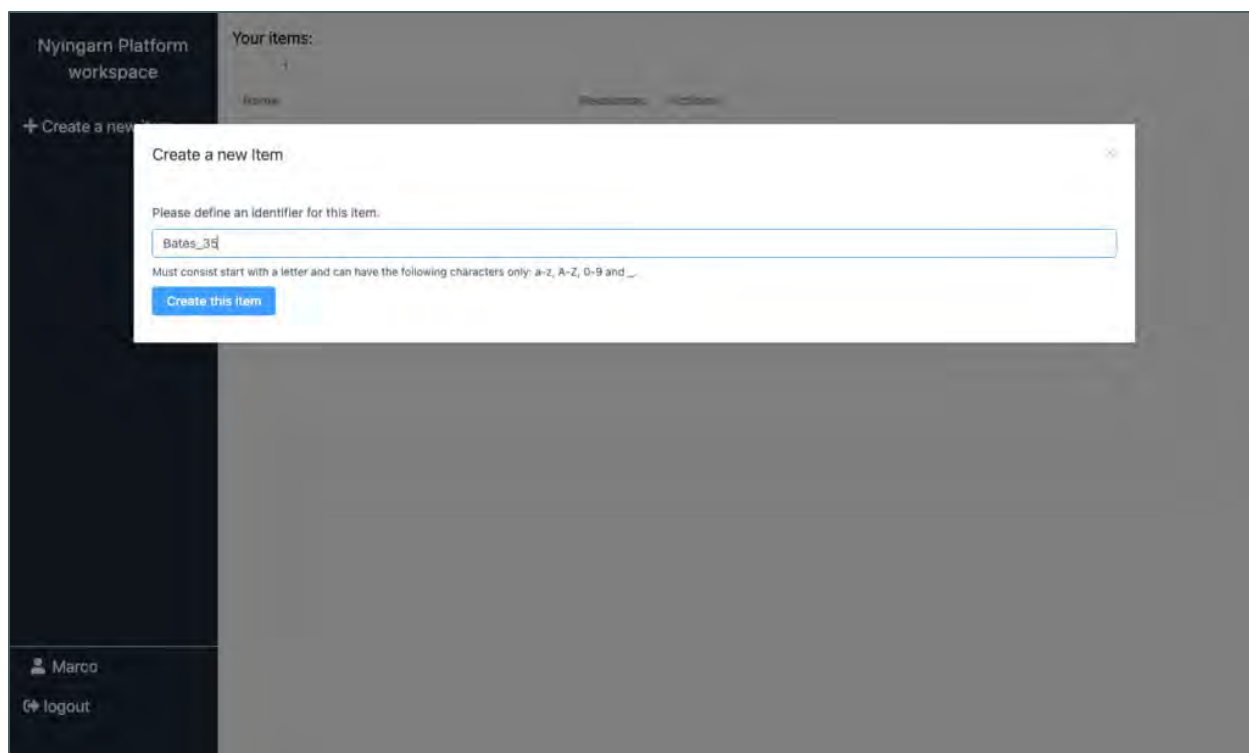When you log in to the workspace, you will be presented with an empty dashboard.
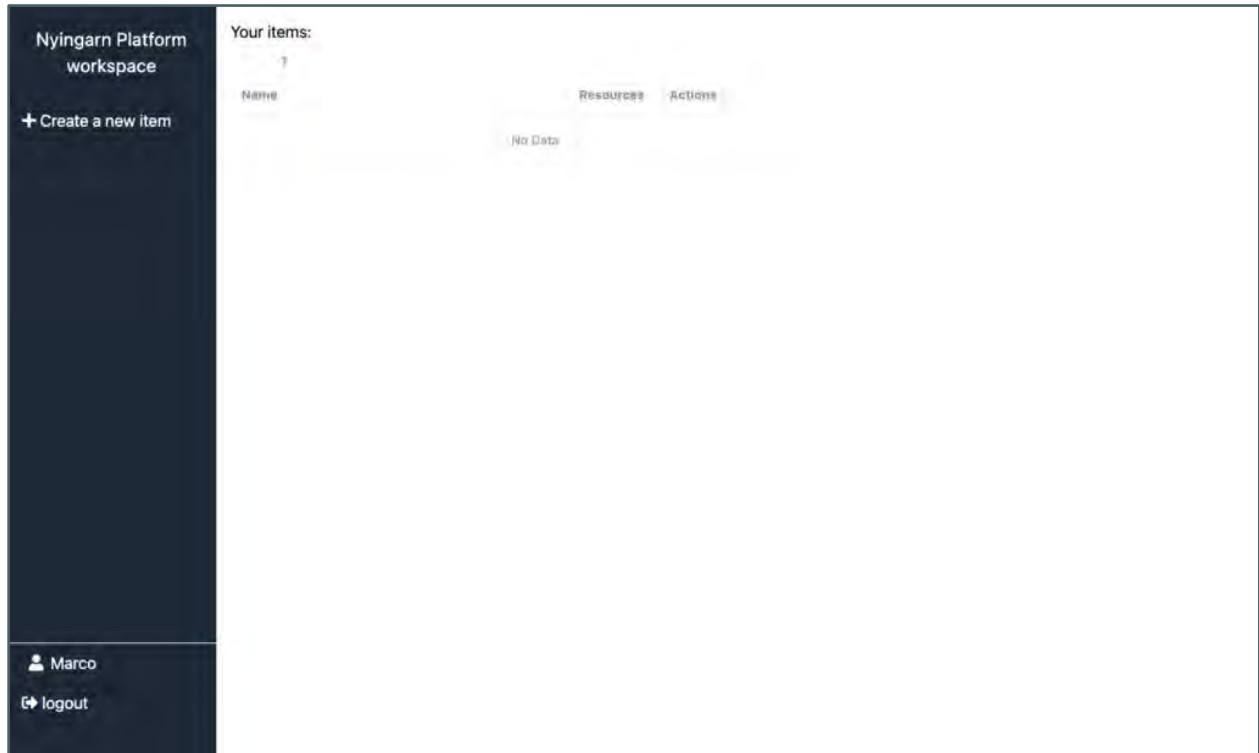
Get started by selecting *+Create a new item* in the sidebar. When naming your item you must meet the following requirements:

- the name must start with a letter
- the name can have any sequence of uppercase or lowercase letters, digits (0-9) and underscore (_)
- the name cannot have a dash in it (-)
- The name must be unique within the Workspace – the Workspace will prevent you from creating a name that already exists.

This name will be the identifier for your item. Some examples include Bates34 or SLNSW_FL814, which relate to the collector's name or the holding institution naming system.

2.2 Creating a collection

When you log in to the Workspace, you will be presented with your dashboard.

Get started by selecting *Create collection* in the sidebar. When naming your collection, you must meet the following requirements:

- the name must start with a letter
- the name can have any sequence of uppercase or lower case characters, digits (0-9) and underscore (_)
- the name cannot have a dash in it (-)
- the name must be unique within the Workspace – the Workspace will prevent you from creating a name that already exists.
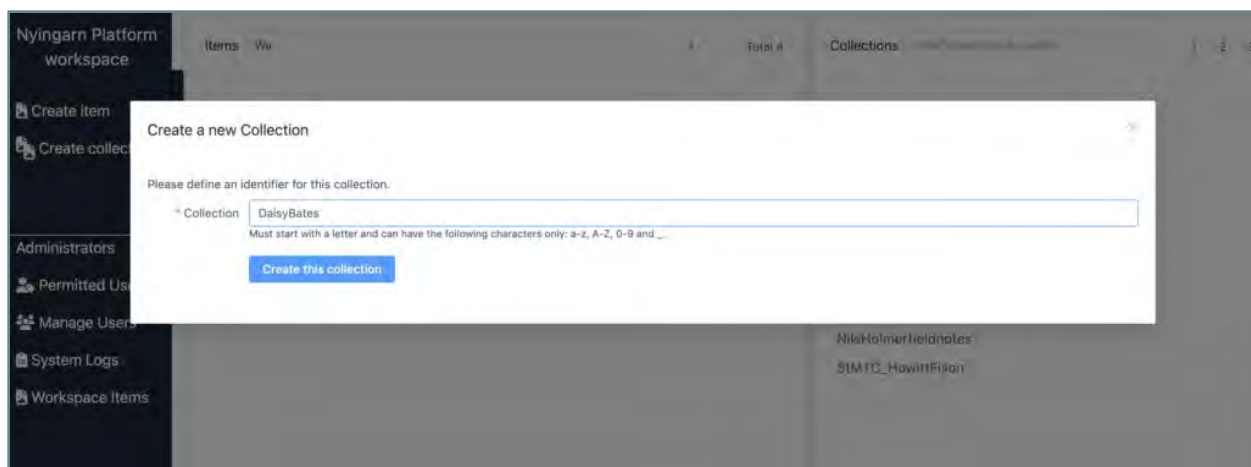
Items can be added to your collection using the *Associate Collections and Items* tab. Click the blue 'link' button on the right hand side to associate an item with your collection. To remove items, click the red 'unlink' button.
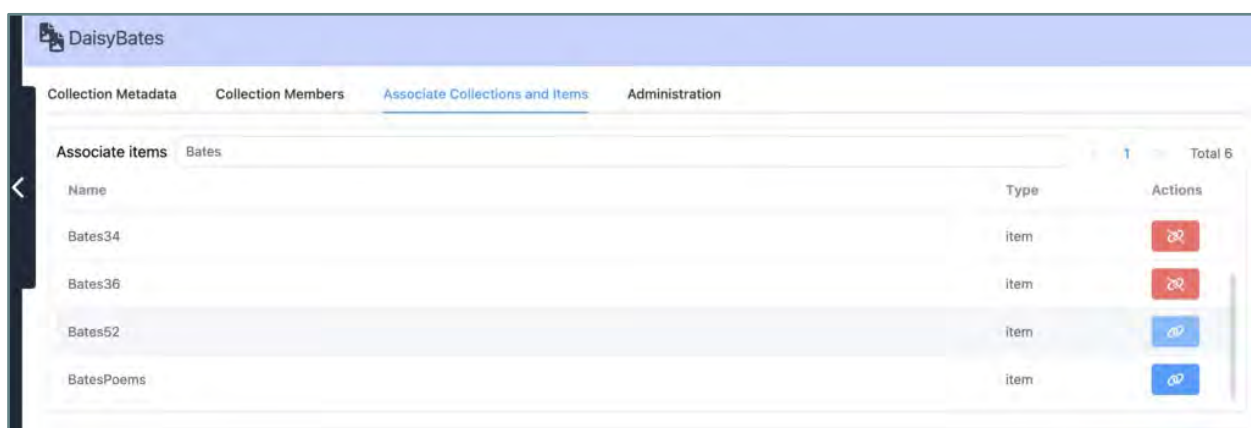
2.3 Preparing your manuscript to upload

To upload a manuscript to the Nyingarn Workspace, you will need to have each manuscript page as a separate image file (JPEG or TIFF). PDF documents cannot be uploaded as manuscript images.

Converting PDFs to image formats

If you are working with a PDF file, convert your PDF to image formats. There are different ways you can do this. Options include using the built in conversion in Adobe Acrobat, or using an online conversion tool.

Splitting images

Sometimes manuscript documents contain double page spreads. It can be helpful to split these scans into individual pages so that each page of transcription in Nyingarn reflects a single page of the original manuscript.
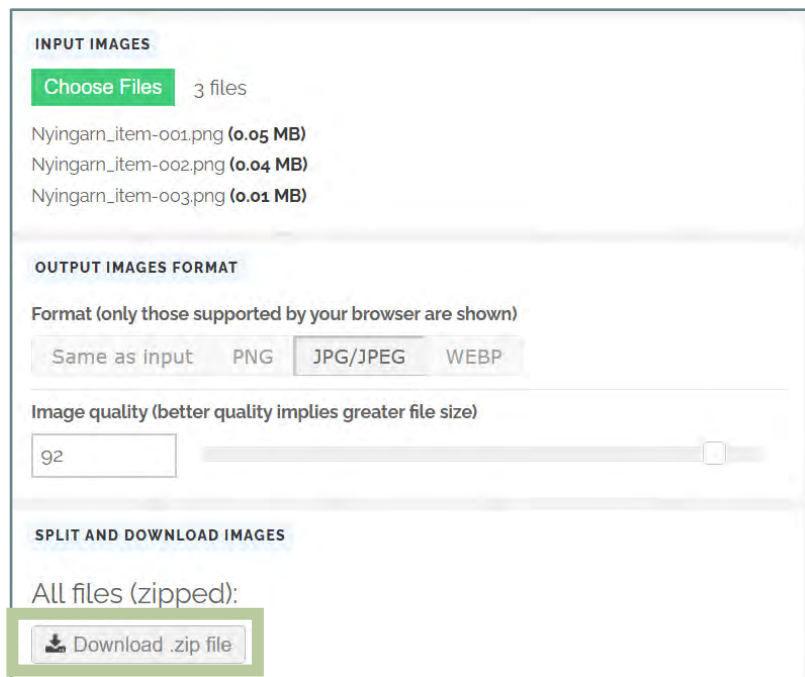
One way to do this is to use an online tool such as Pine Tools Bulk Image Split and split each page. Each new page (half of the original scanned double spread) can either be sequenced as -001, -002 etc or as -001A, -001B.

To do this:

> **Step 1.** Open https://pinetools.com/bulk-batch-split-image

> **Step 2.** Drag and drop your JPEG files into Pinetools or 'Choose files' under the 'input images' field

> **Step 3.** Click 'download .zip file'

> **Step 4.** Open the zip folder in your downloads to extract the files then save onto your computer.

Naming your manuscript images

You may need to rename your manuscript image files before uploading them to the Nyingarn Workspace. Your image files must be named consistently with your item identifier to be ingested.

The screenshot below shows files and folders from the Daisy Bates Collection, University of Adelaide Library. The collection is organised into 13 Chapters of 66 Folios, and as digital files looks like this:
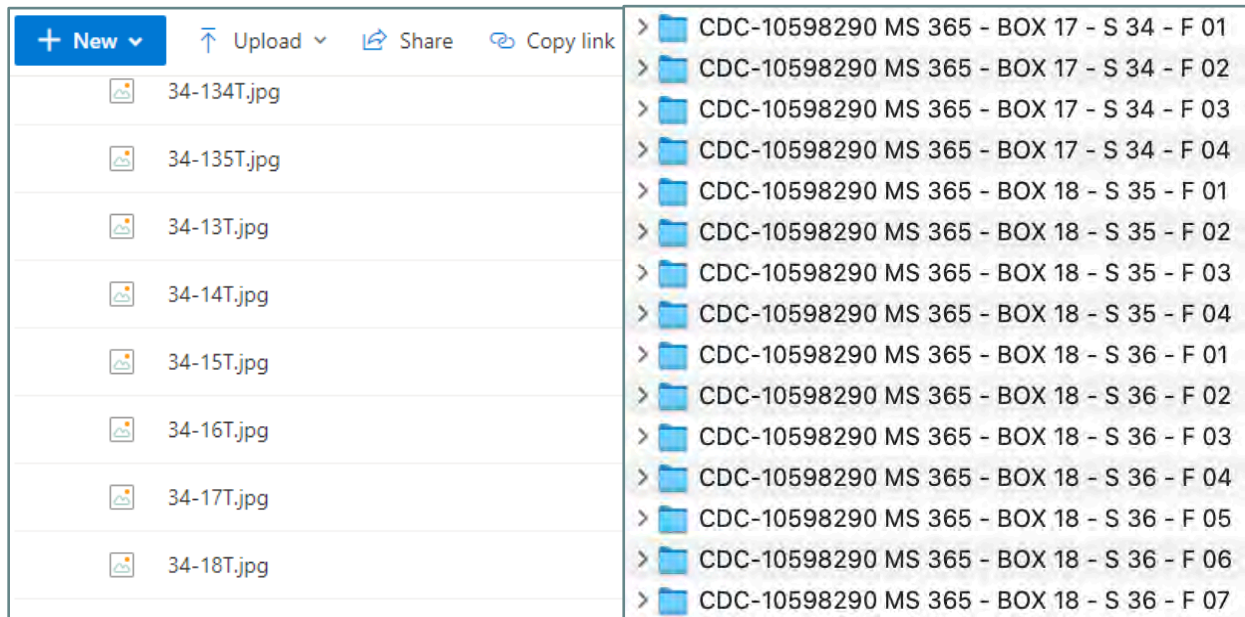
Naming these files for the Nyingarn Workspace:

- The item can be the whole manuscript or a section of the manuscript.
- For large collections like Daisy Bates, the item was identified as a section. For example, Bates34, Bates35, Bates36 and so on. Items were then created with these names.
- The manuscript image pages were then renamed following this convention: Bates34-001.jpg, Bates34-002.jpg, Bates34-003.jpg.
- Looking at the above left-hand screenshot, 34-15T.jpg becomes Bates34-015.jpg.
- Small manuscripts (less than 10 pages) can be named MSS12-1.tiff, MSS12-2.tiff, without the additional 0.
- For larger manuscripts (more than 10 pages) we recommend adding the extra 0/00 numerals like this MSS12-001.tiff, MSS12-002.tiff … MSS12.010.tiff.

NOTE: The reason we need to add 0s to the end of the image file names is to allow the computer to count the pages and order them correctly in Nyingarn. As an example, if you have a manuscript that is over 100 pages long, you need to allow the computer to count to three digits – so you will need to name your files 'example-001'. If you have a manuscript that is 1000 pages or more then you will need to add an extra zero to the file name 'example-0001. If the manuscript is less than 100 pages, then you only need to add one 0 - 'example-01'.
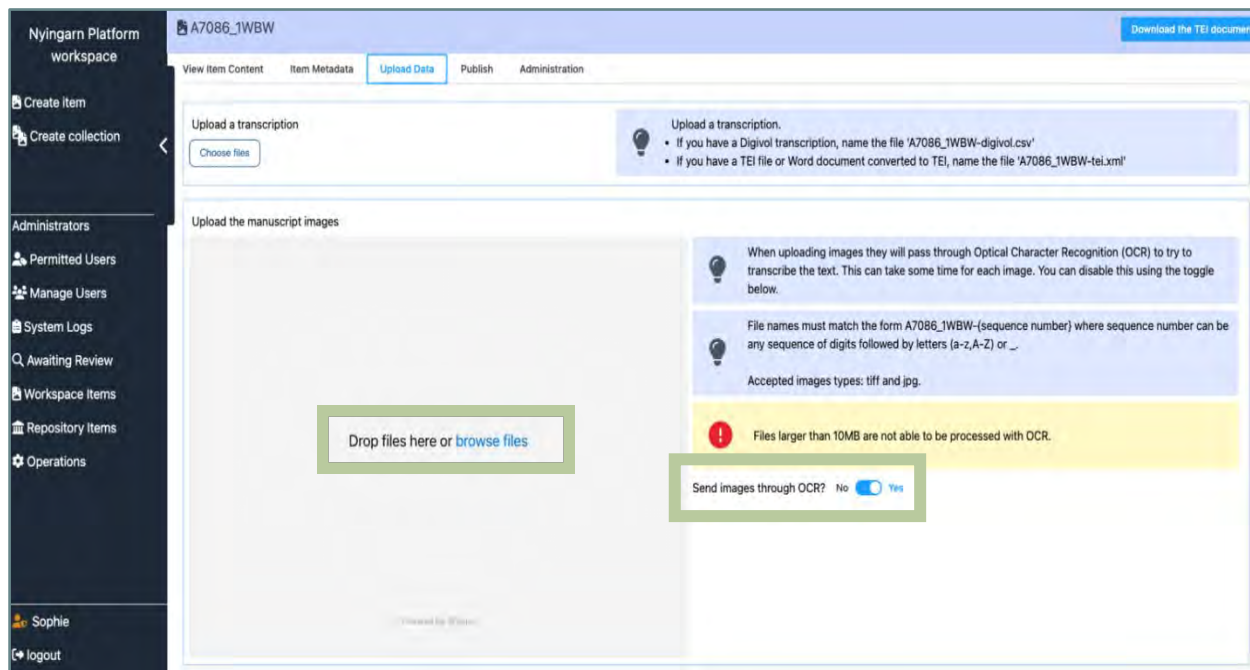
2.4 Uploading manuscript images for transcription

Use this option in the Workspace when you have manuscript images that require Optical Character Recognition (OCR) transcription.

Step 1. After logging in to the Workspace, click into an item on your dashboard or create a new item

Step 2. Click on the *Upload Data* tab

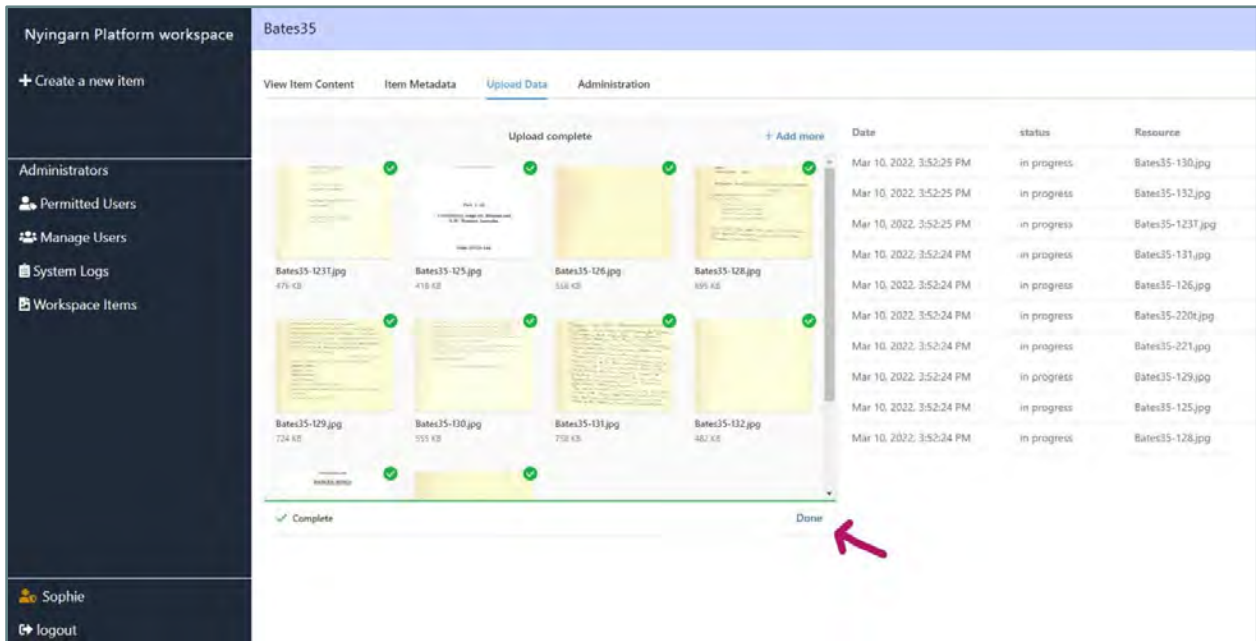**Step 3.** Then click *browse files* or drop files into the grey box

Notice the green ticks, as the Workspace accepts each manuscript file. Each of these resources is now ready for automated OCR. Click *Done* to start the processing. To send images through OCR, make sure the *Send images through OCR* option is toggled to 'yes' (default).

OCR may take a few minutes, depending on how many files you are uploading and who else is working in the Workspace. If in doubt, refresh your browser. When ready, you can view manuscript thumbnails in the *View Item Content* tab.
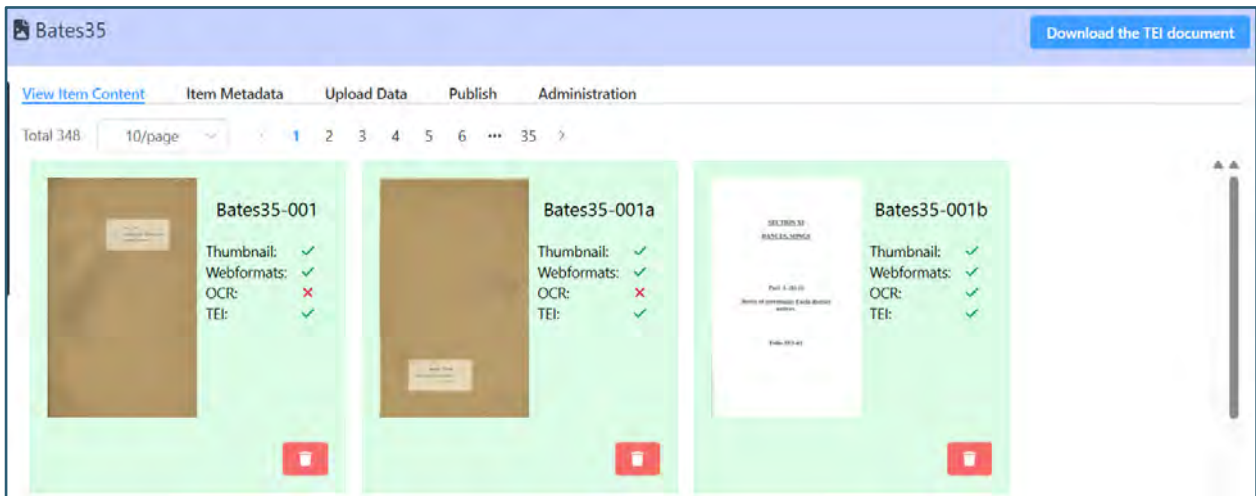
3. Uploading pre-existing transcriptions

3.1 Uploading a transcription

The Nyingarn Workspace can ingest your DigiVol or FromthePage transcription work.

DigiVol

Once your transcription and validation work is complete, navigate to the *Expedition Administration* screen. Click the *Tools* button to view the export options.

Select *Export all tasks*. Then choose single de-normalised CSV file.

Name your CSV file consistently with your Nyingarn item name, e.g. Bates35-digivol.csv.

Note: When uploading image files for transcription in DigiVol, make sure they are also named consistently with your Nyingarn item name. This allows the Workspace to recognise each page of transcription and match it up with the corresponding manuscript image.

Upload your DigiVol CSV file to the *Upload a Transcription* box in the Nyingarn Workspace. When the Workspace receives the DigiVol CSV file it extracts the transcriptions for each image and creates a file for each.
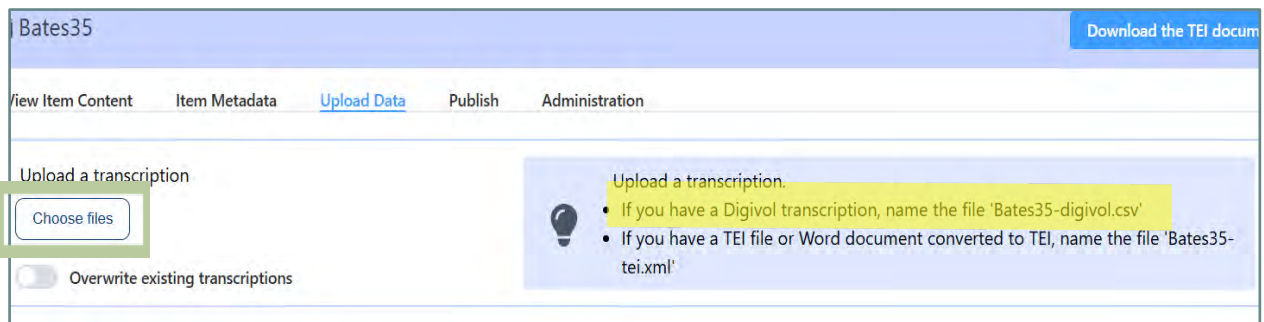
Once the CSV file has been processed you should see placeholders for each resource defined in the file in the *View Item Content* tab.
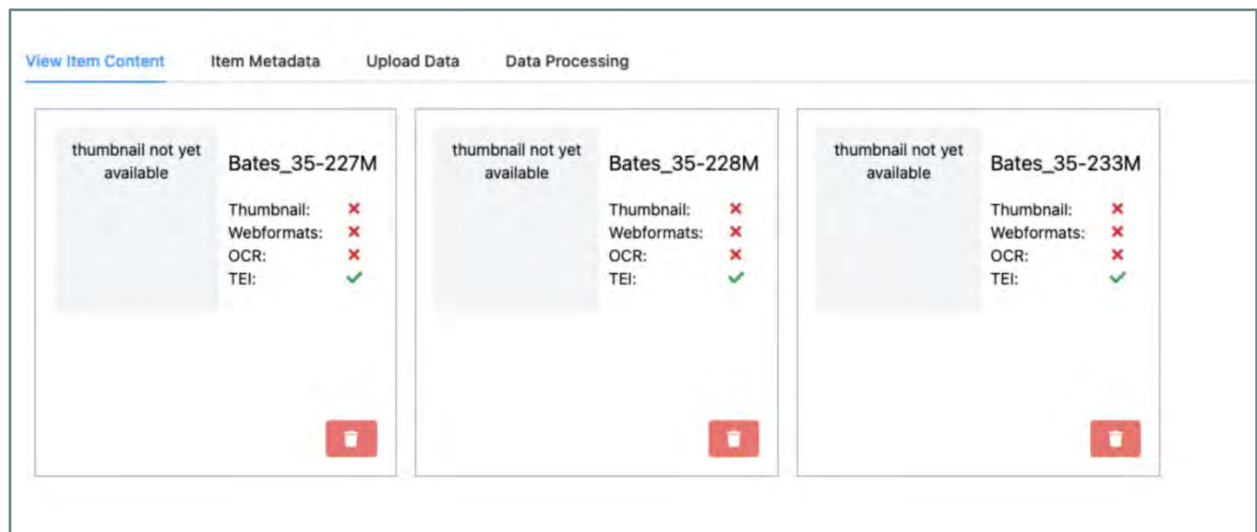
At this point, you can upload the manuscript image files. When you upload the image files, toggle *Send images through OCR* to 'no'. As these are ingested into the Workspace they will go through processing stages to produce thumbnails and various formats suitable for display on the web. Once complete, the *Item Content* tab will look like this:

Notice the green tick for 'Thumbnail', 'Webformats' and 'TEI' for each image. Each of these resources is now ready for further work inside the Workspace. Also notice that OCR is not ticked. When ingesting transcriptions from DigiVol the OCR step is skipped.

FromthePage

Once your transcription work is complete, use the export tab to view your individual works.

Navigate to the work you want to export from your work title list. For Nyingarn, you will need to download the TEI version of your transcription.

The XML file will appear on your screen. Right click to save and name.

The process for uploading FromThePage data is the same as uploading DIGIVOL data except that the file is named differently. In this case, the file must be named "{name of your item}-tei.xml".
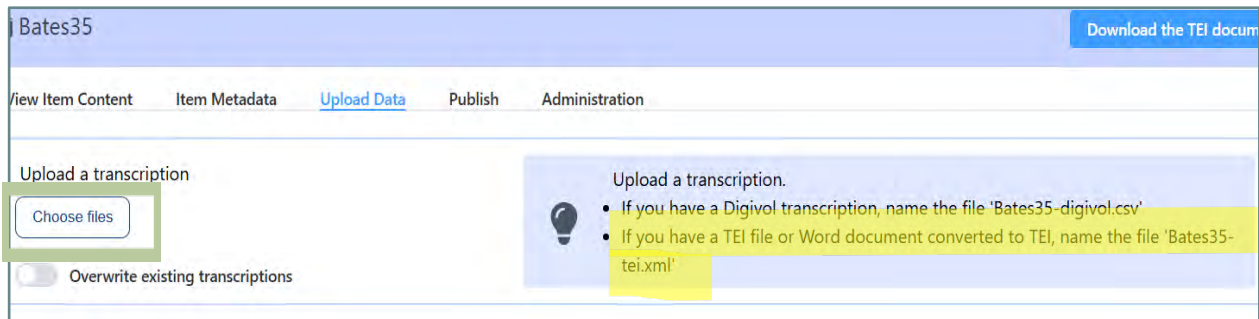
3.2 Converting a Word document to TEI

If you have a transcription done in Microsoft Word or similar, these existing transcription text files will need to be converted to TEI XML format before uploading them to the Nyingarn Workspace. The Workspace will accept Microsoft Word transcriptions modified for TEI and transformed into TEI.XML through TEIgarage, an online conversion service. Microsoft Word documents may contain headings and page breaks obstructing the conversion process. These unnecessary elements need to be removed, and the page numbers appropriately styled to be recognised by the code in the Workspace.

Before starting, decide on the manuscript item name. Remember that in Nyingarn, an item name must be a unique identifier and can contain letters, numbers, and underscores. The page number sequencing then follows the item name. For example Bates34-001; Bates34-002; Bates34-003.

Naming your pages in sequence

Name each page of your transcription according to the Nyingarn item name and sequence number. For example, on the first page of the transcription you would type Bates34-001 at the top or bottom of the page. This page of transcription should correspond to the same page of the original manuscript (i.e. page 1).

Creating a style of page numbers

Step 1. In Word, click the *Styles* group in the Home toolbar.

Step 2. Click the *A+* at the bottom of the tool window to create a new style.
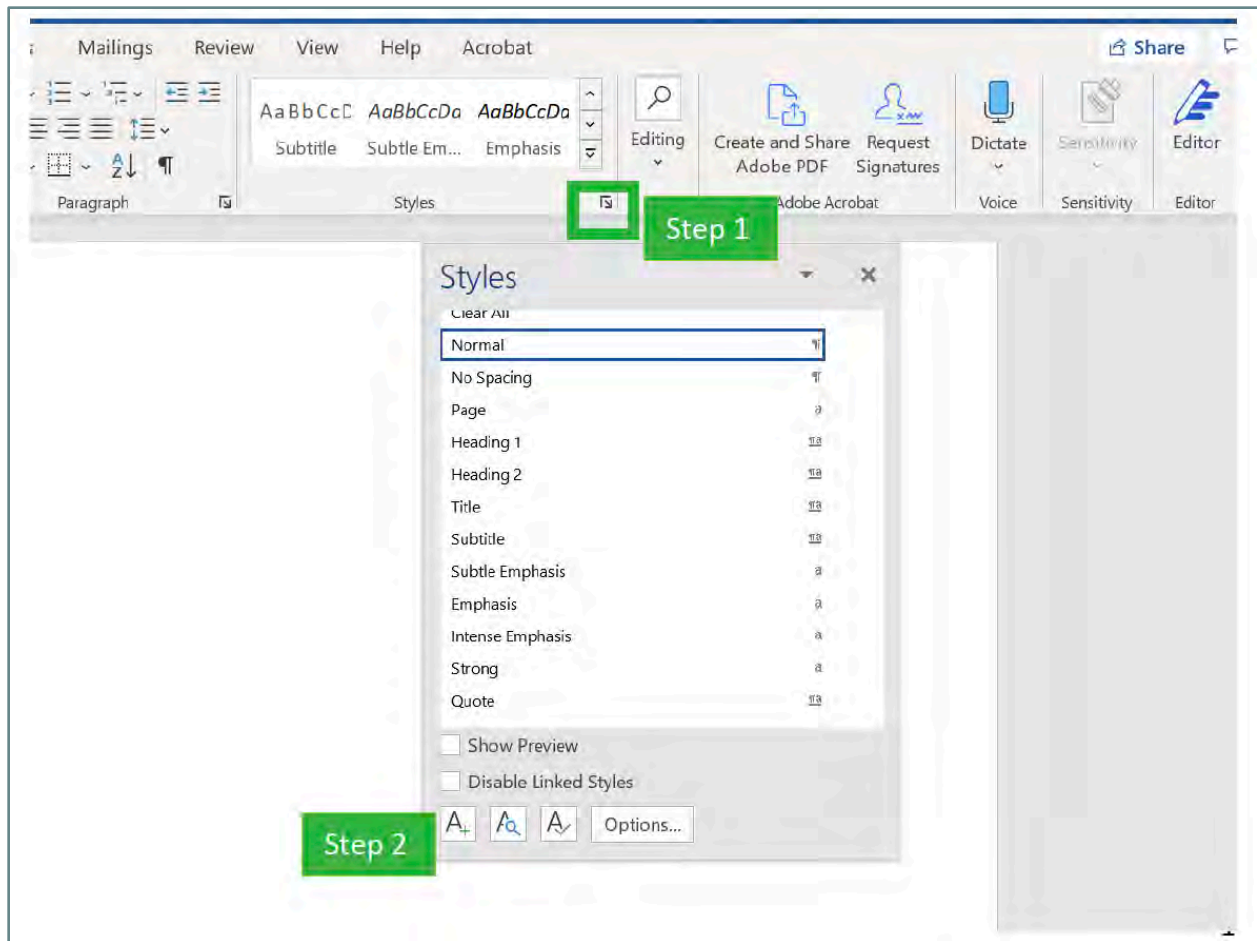
Step 3. Name the style 'Page'.

Step 4. For style type choose 'Character' in the dropdown menu.

Step 5. Choose a Style colour. Choosing a colour other than black will help you to recognise the change in your document.

Page is now a standard style in your document.
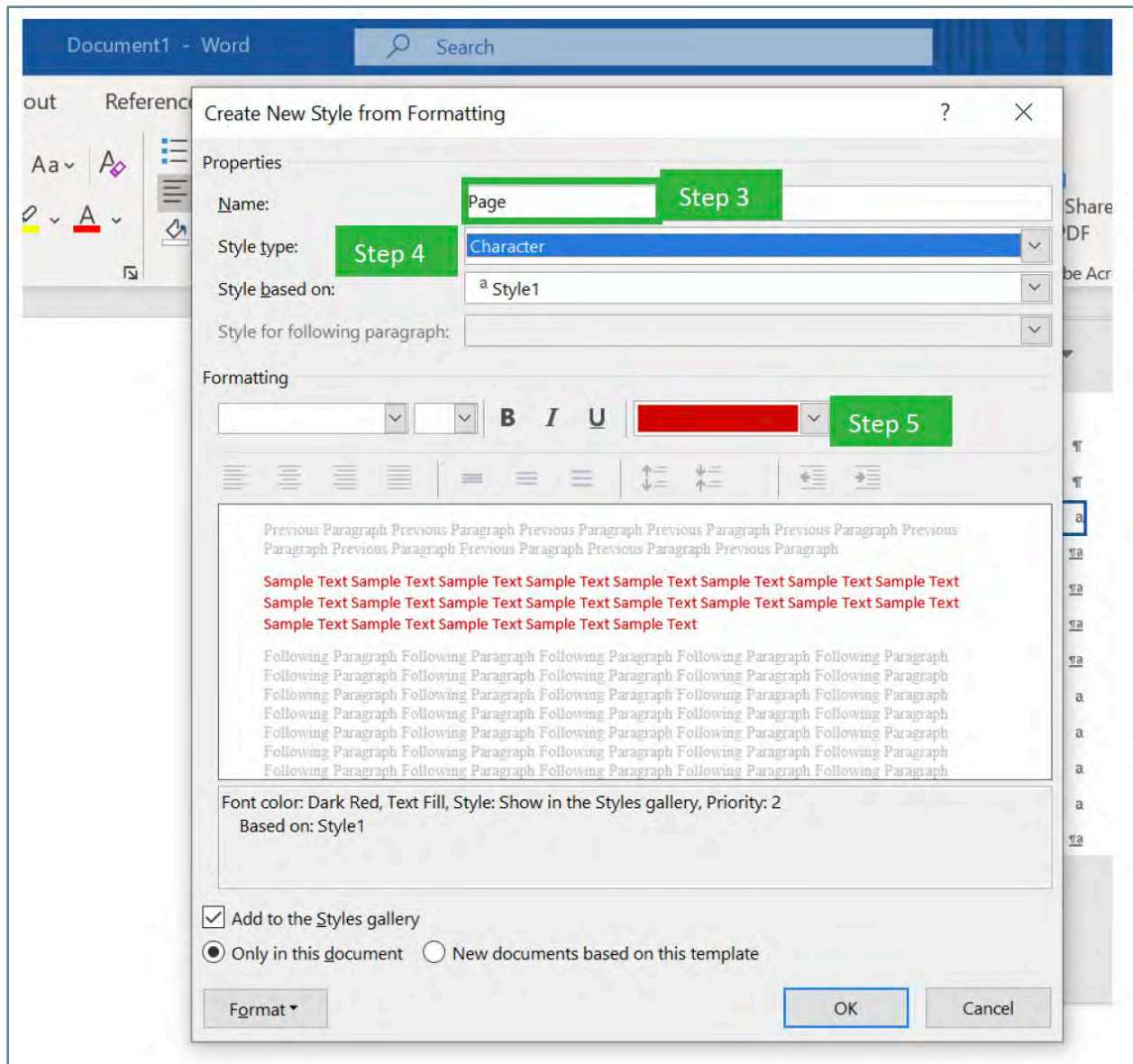
Apply Page style to every page name (e.g. Bates34-001) in the document

Step 6. The find and replace function is helpful for bulk changes. Type 'Bates34-???' into the *Find what:* box. The ??? denote wild fields and will help you to find the entire page number.

Note: Make sure *use wildcards* is ticked.

Step 7. Click the cursor in the *Replace with:* field, then click *More* to expand the Find and Replace options if they are not already displayed.
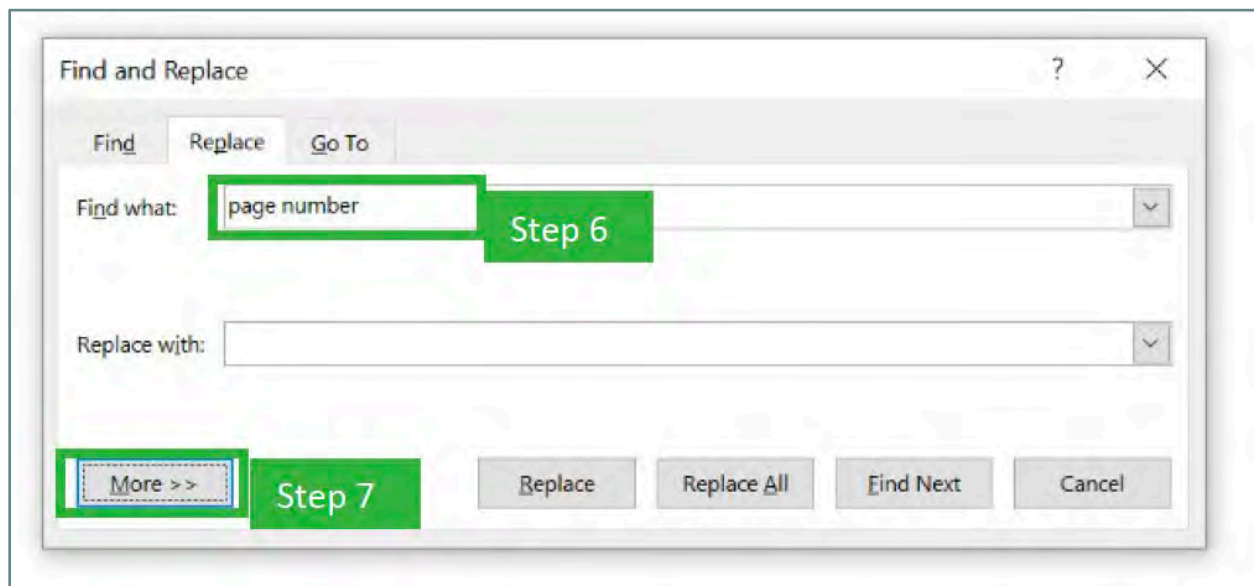
Step 8. Next click the *Format* button.

Step 9. Click the *Style* option, and choose style 'Page'. Click 'Replace All'.
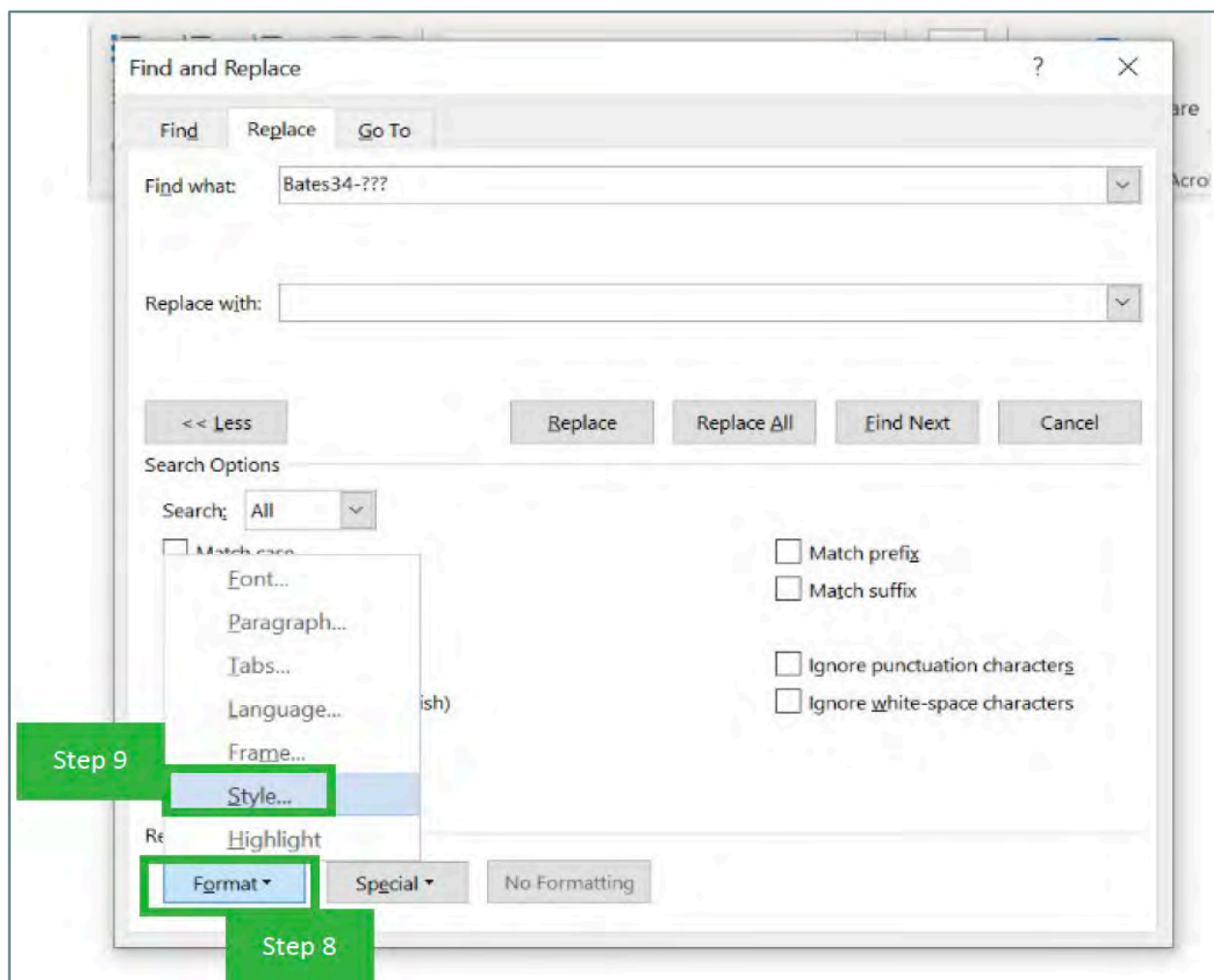
Now that the page naming/numbering is correct and styled with Page style, page breaks should be removed.

Remove Page Breaks

Do these steps if you know there are page breaks in your manuscript, or if you aren't sure.

Step 10. Using the Find and Replace function, in the *Find what:* field click 'Special'.

Step 11. Choose 'Manual Page Break'. This will add the symbol ^m (see screenshot below). Nothing is needed in the *Replace with:* section, so leave it blank. Click 'Replace All'.
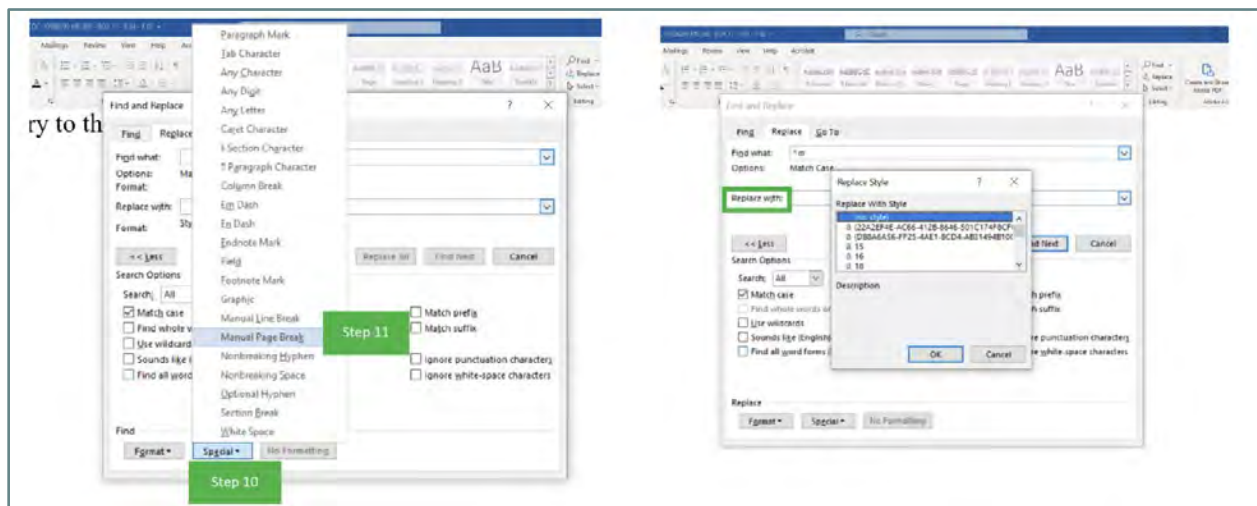
Final Steps – Save and convert the document

Step 12. Save the Microsoft Word .docx file using the naming convention for the Nyingarn Workspace e.g. Bates34-tei.docx.

Step 13. Convert the .docx file to TEI XML using TEIgarage: https://teigarage.tei-c.org/#. TEIgarage will ask you to select the type of document you want to convert.

Step 14. Choose *Documents*. Choose Convert from: Microsoft (.docx), Convert to: TEI P5 XML Document.

Step 15. The next window will ask you to select the file for conversion. Click the *Choose File* button to browse, find, and upload your file.

Step 16. The file will automatically download ready to be ingested into the Nyingarn Workspace. The document should be named e.g. Bates34-tei.xml.

4. Editing and enhancing your transcription

4.1 Editing your transcription

You can edit and enhance your transcription by typing into the transcription textbox. Click into a thumbnail in the *View Item Content* tab to begin.

The manuscript page is shown on the left-hand side and the transcription text box on the right. The transcription box has two tabs: WRITE and PREVIEW. In the WRITE tab, the user can edit and enhance the text. The TEI XML markup is colour-coded around the text. In the PREVIEW tab, the user sees only the transcription text.
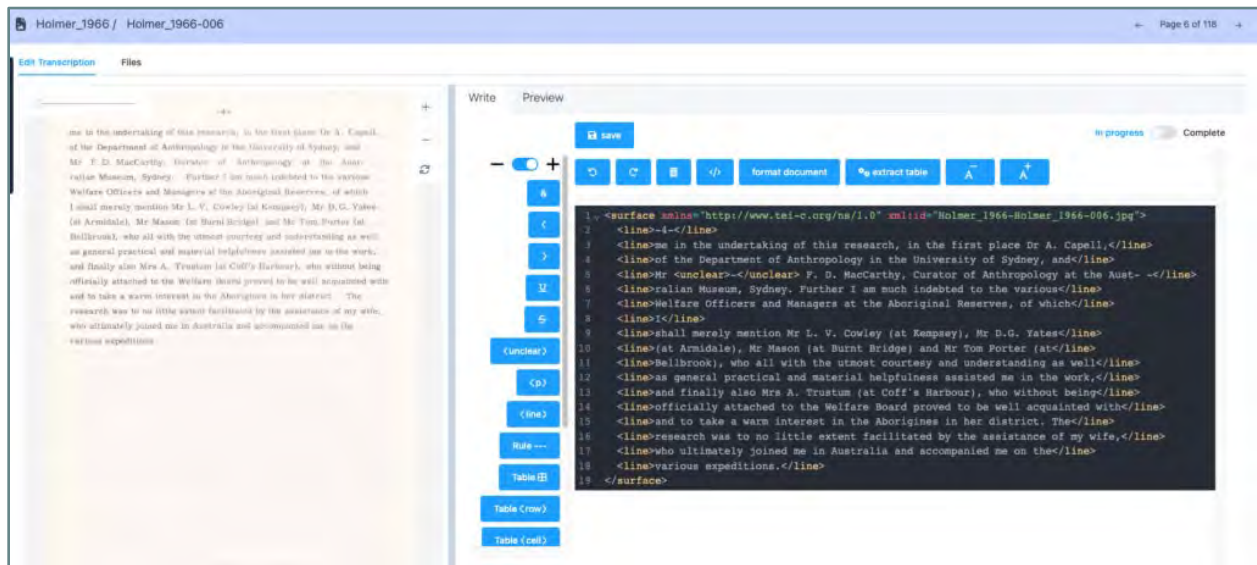
To edit the text, click in the text box and add or remove content.

To add markup options, highlight the text and click the relevant blue markup button. The buttons sit in the centre of your screen between the manuscript and transcription windows.

https://schwa.dk/filer/ipacharpick/ provides many useful characters and symbols. The symbols must be Unicode.

To remove markup, change the -/+ slide button to – (buttons will turn red), highlight the text and click the relevant red markup button.
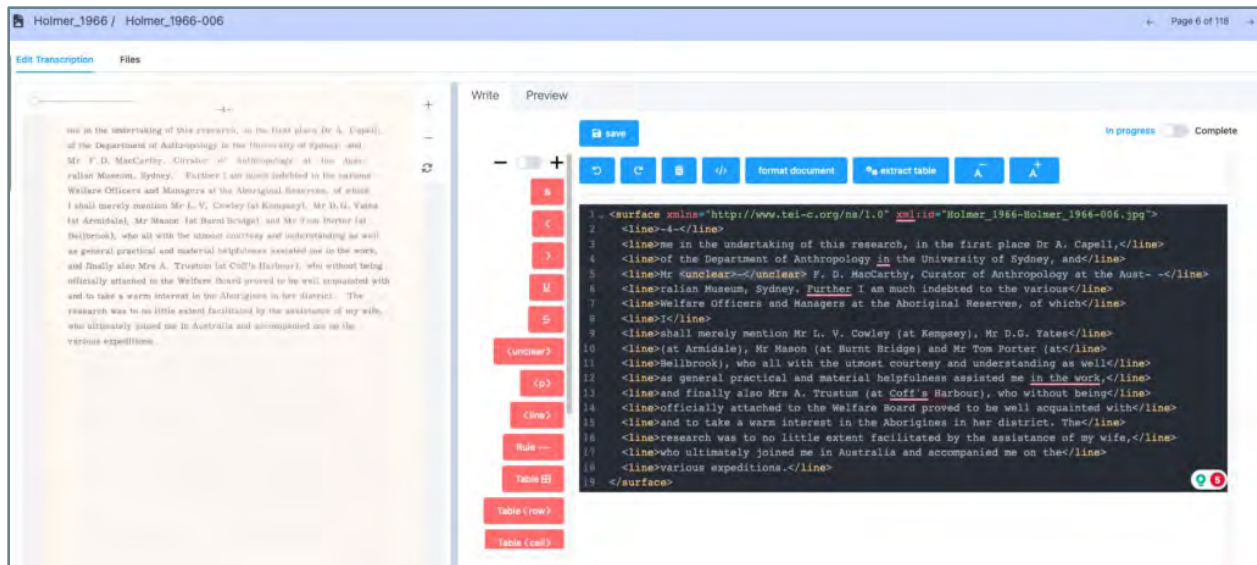
4.2 Enhancing your transcription

Important information like place names, people, community/tribe, and language can be marked up within the document. This manuscript example is written in Italian and a Western Australian Indigenous language. English translations have been added and marked as additional English language information using this code: '<term XML: lang="en" </term>.
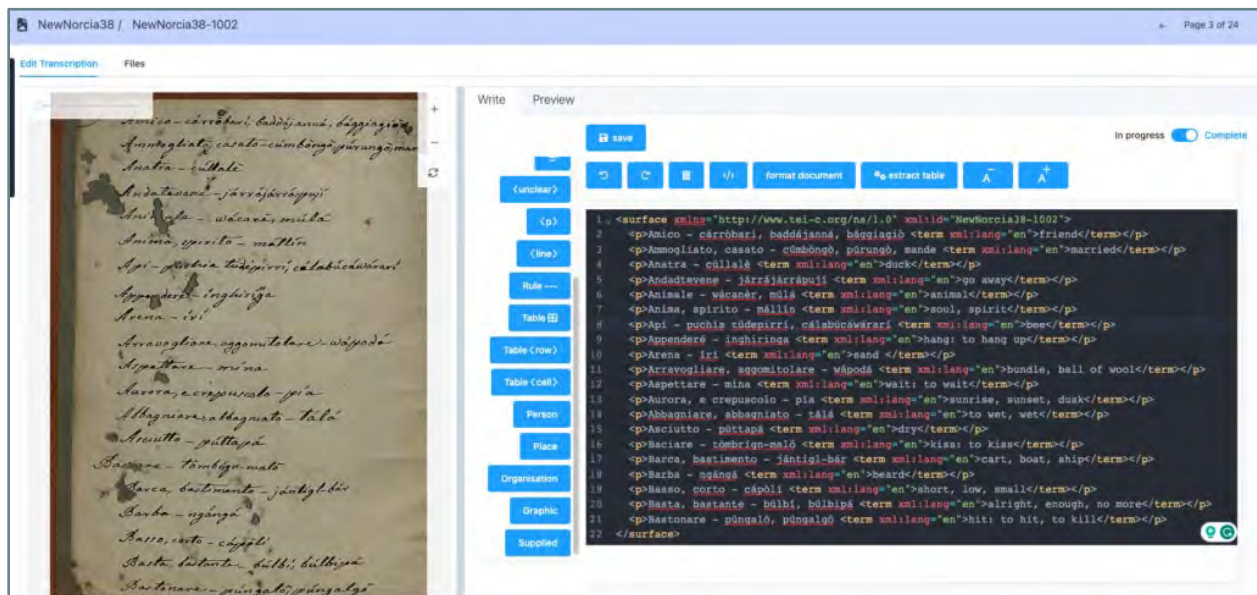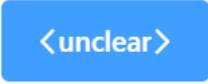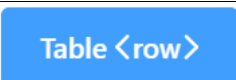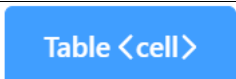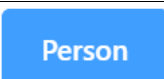
The *supplied* markup button could also be used. The supplied button can be used to highlight parts of the transcription not in the original manuscript, such as a translation:
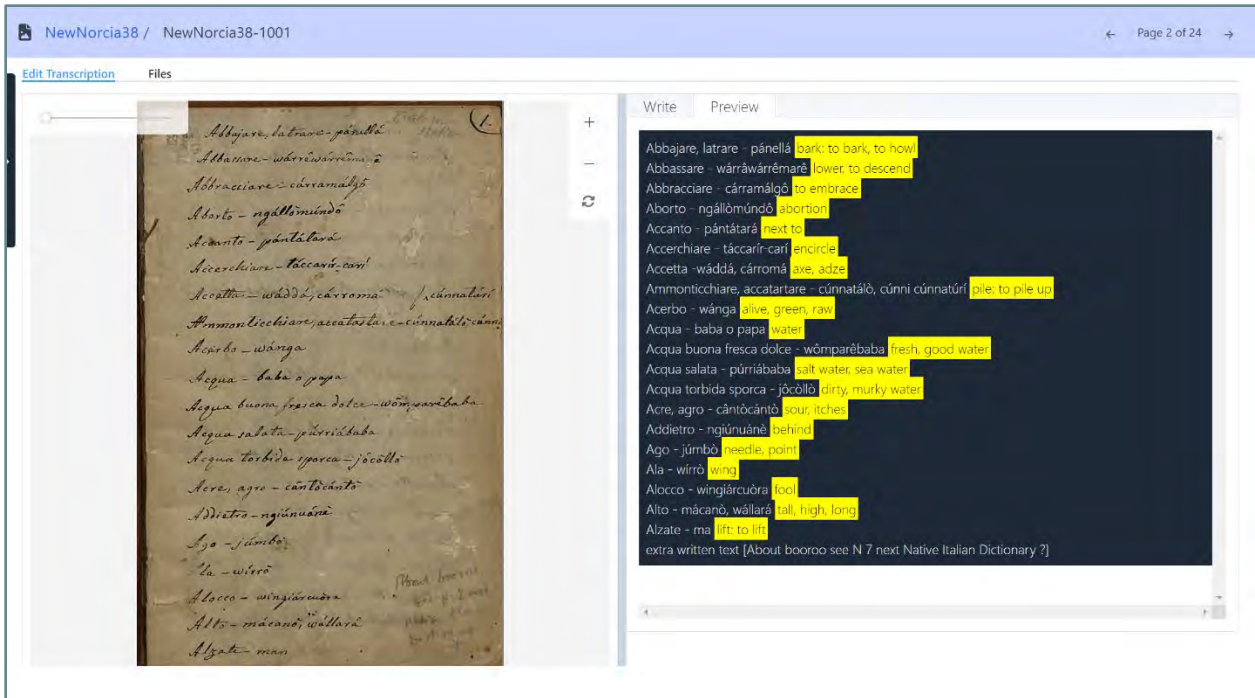
Another example is to supply the names of people referred to by their initials:
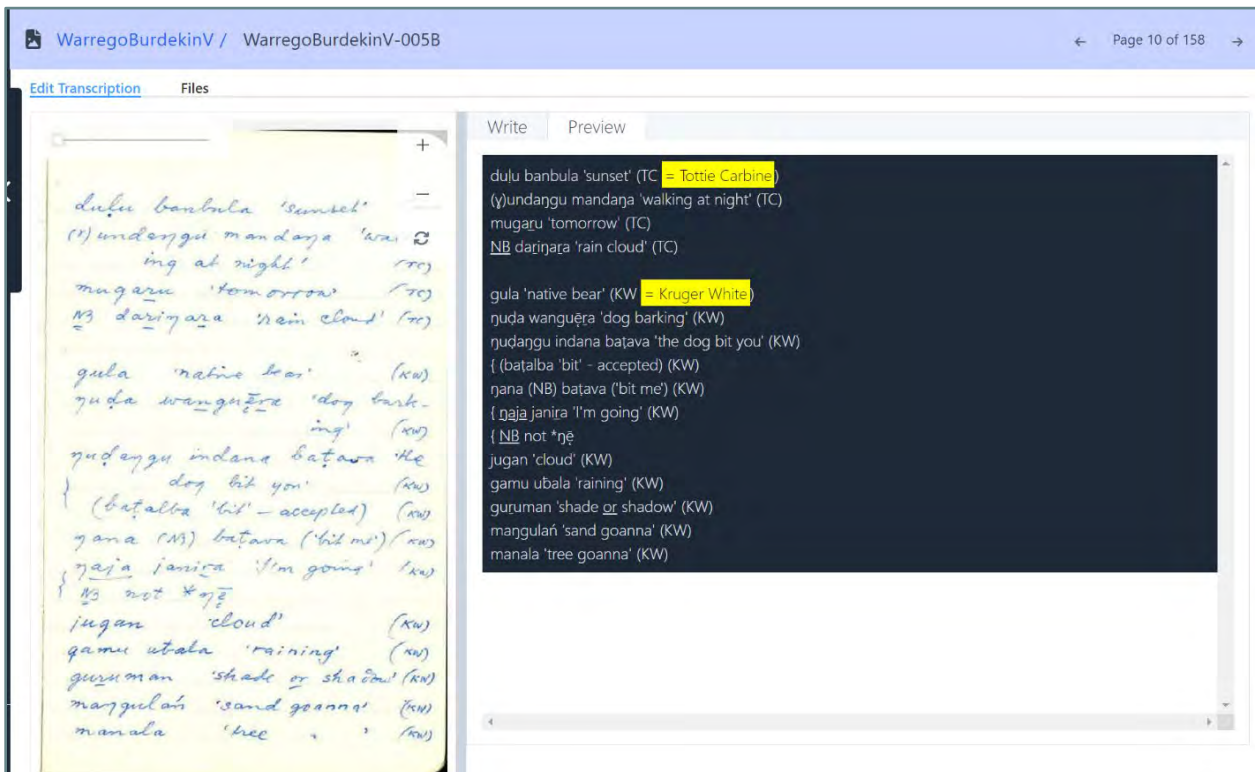
4.3 The TEI markup buttons

&	Inputs an ampersand '&' into the transcription. This cannot be done using your keyboard as the ampersand is used within the coding language.
<	Inputs a left chevron '<' into the transcription. This cannot be done using your keyboard as the chevron is used within the coding language.
>	Inputs a right chevron '>' into the transcription. This cannot be done using your keyboard as the chevron is used within the coding language.
U	Underlines selected text.
S	Strikes through selected text.
\<unclear\>	Used when a word or section in the manuscript is unclear. The content contained within this markup button is the transcriber's best guess. Unclear text is underlined in red.
\<p\>	Adds a paragraph break to the transcription. This helps to visually format the transcription for easier reading.
\<line\>	Adds a line to the transcription. This ensures that the transcription does not look like a single block of text.
Rule ---	Used to input a horizontal line. This is a common feature in some historical manuscripts.
Table ⊞	Starts the process of making a table — see section 4.4 for more information.
Table \<row\>	Adds rows to tables — see section 4.4 for more information.
Table \<cell\>	Adds cells to table rows — see section 4.4 for more information.
Person	Used to tag a person's name in the manuscript. This allows users to search for a particular person in manuscripts.

Place	Used to tag a place name in the manuscript. This allows users to search for a location in manuscripts.
Organisation	Used to tag an 'organisation' in the manuscript. An organisation can be a community or language group, tribe, clan etc. This allows users to search for a group of people in manuscripts.
Graphic	Used to show the reader that an image exists in the original manuscript (as Nyingarn cannot draw its own images).
Supplied	Used to indicate 'supplied' text that does not appear in the original manuscript. Supplied text is highlighted in yellow.

4.4 Creating tables

Tables can be added to your transcription as a way of organising the text. Here is an example of an automated Optical Character Recognition (OCR) transcription that has not recognised the information in table format.

First, try the extract table blue button. This text isn't technically a table so it may not work.

Next, manually construct a table using the blue markup buttons <table> <row> <cell>.

It is useful to remove the <line> elements before constructing your table. Pictured below, the slide button is - so the buttons are red. Now you can highlight the text and click <line> to remove all the line elements.

This manuscript is a word list with two columns. Markup for one row of the table would look like this:

```
<table>
        <row>
                <cell>column1</cell>
                <cell>column2</cell>
        </row>
</table>.
```
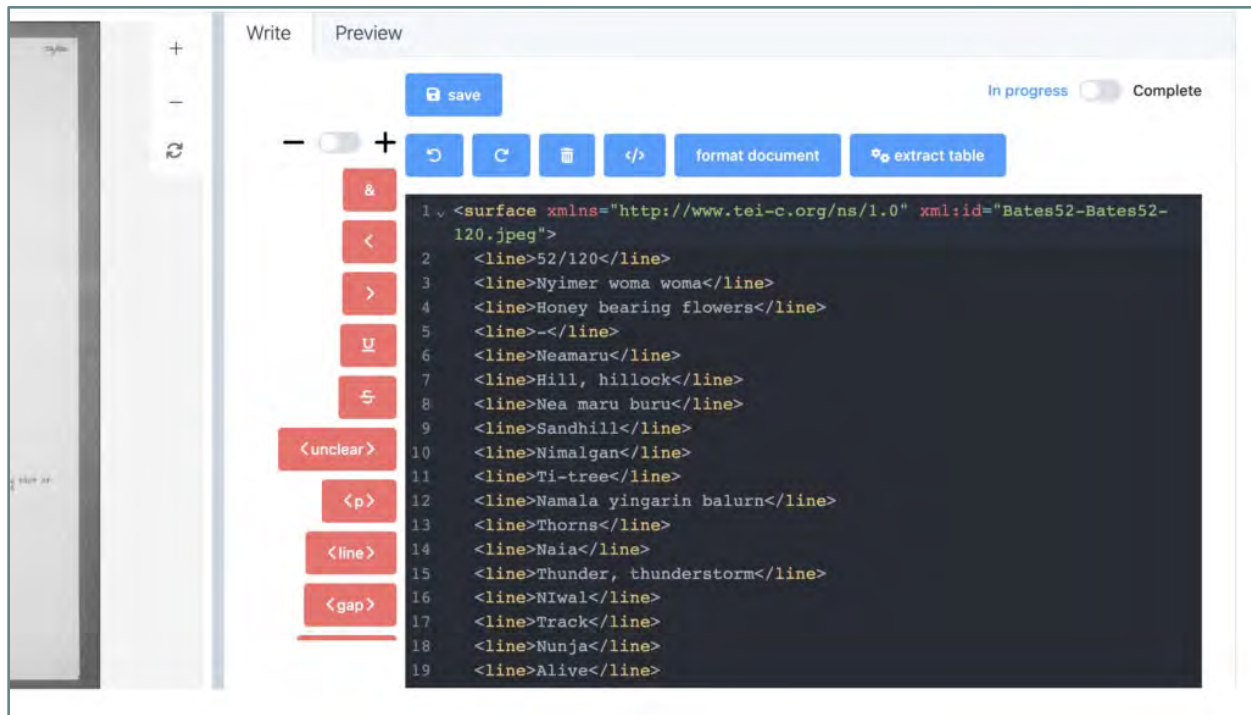
See Workspace example below. The preview tab lets you check the table presentation is correct.
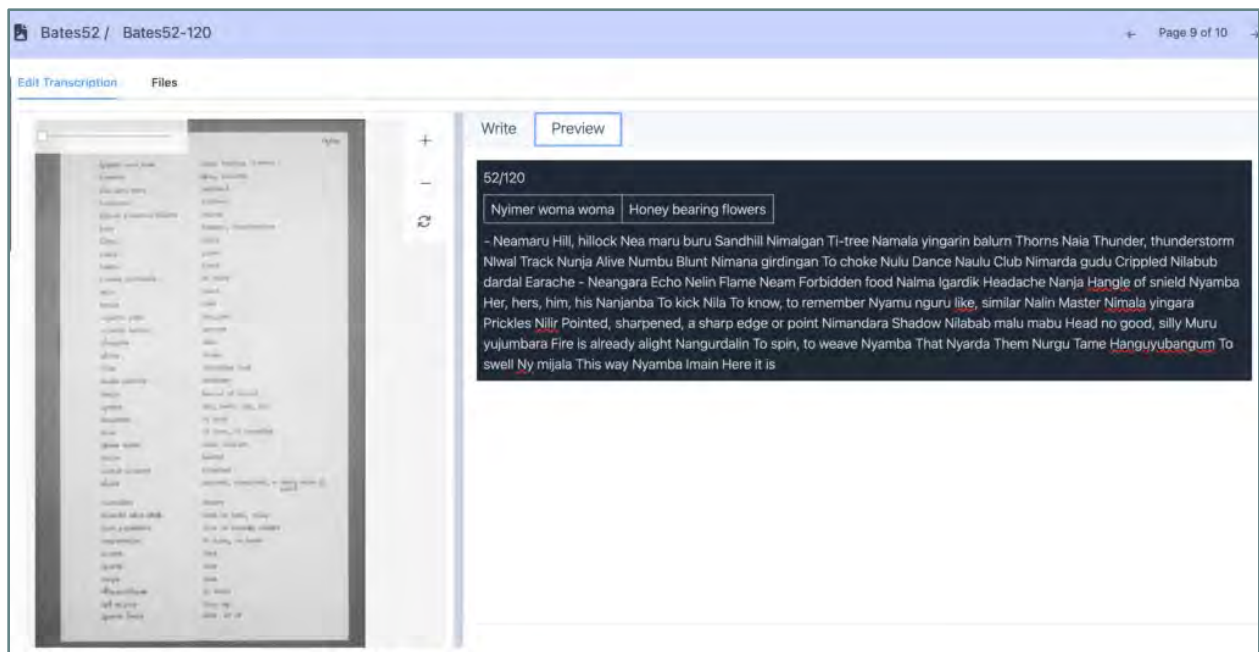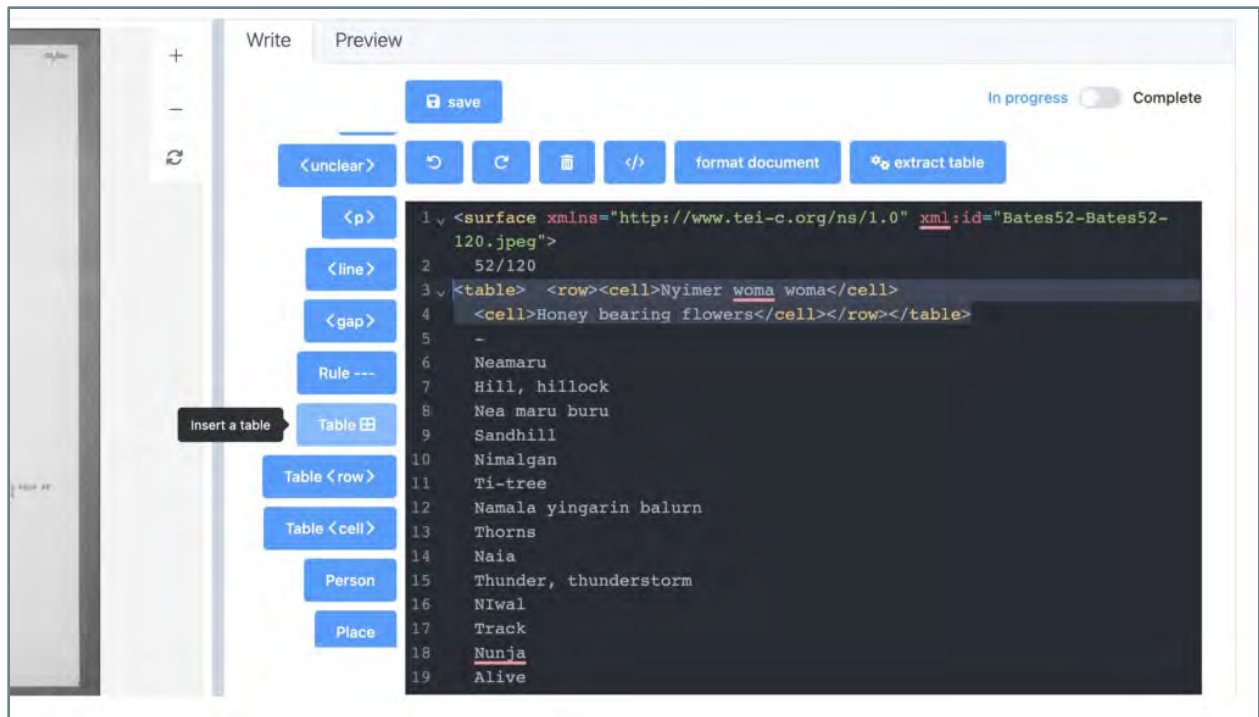
4.5 Using TEI and fixing TEI errors

TEI XML is the coding language that is used for Nyingarn's text transcriptions. The WRITE transcription box allows you to add and remove TEI using buttons. Most of the TEI command buttons produce two parts of code, a beginning and an end. For example, the code **<line>**

signifies the beginning of a new line and the code **</line>** signifies the end of a line. These two codes should sit on either side of the text that they describe.

Dealing with error messages

If there is an error with your TEI code, you will not be able to save the page and will see a red exclamation mark next to the save button. If you hover your mouse over the exclamation mark, it will tell you why and where the error occurs.

Pages with a red background in thumbnail view contain TEI errors. This sometimes means that the TEI document has not yet been produced and the Workspace needs more time to process the page.

'Opening and ending tag mismatch'
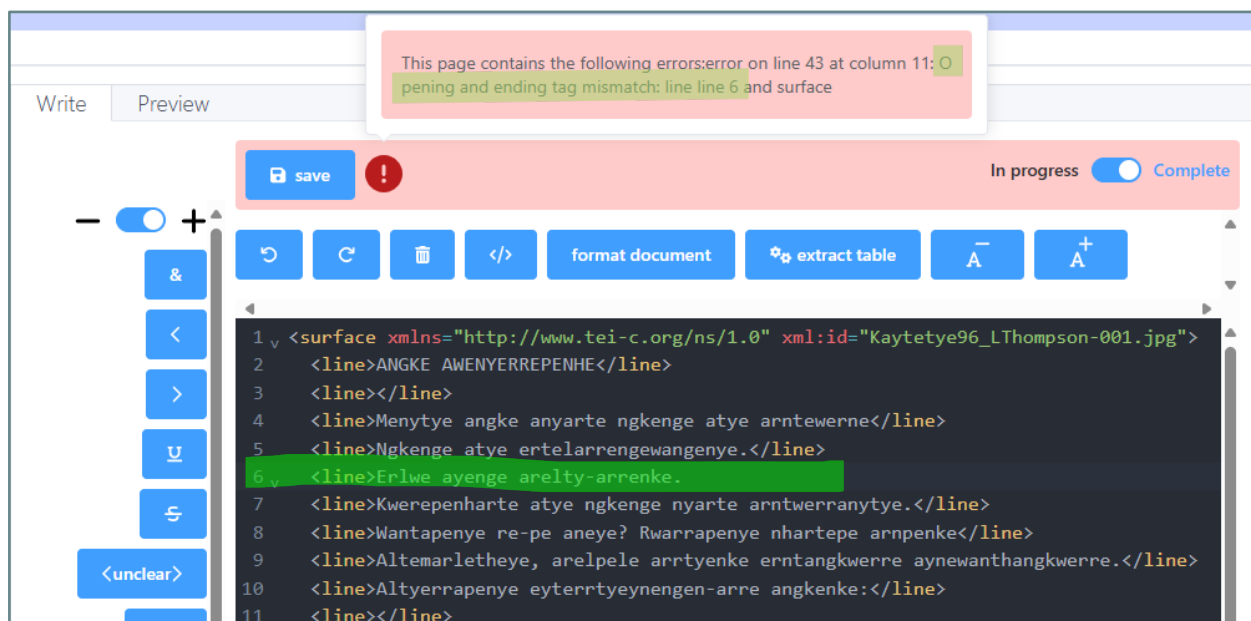
A common TEI error in the Workspace is code appearing without its pair: e.g. **<line>** appearing at the start of the line with no **</line>** at the end of the line.

In the screenshot above, the error message tells us that there is an 'opening and ending tag mismatch' on line 6. To fix the error, we need to add **</line>** to the end of line 6.

'Extra content at the end of the document'

Each page of your transcription needs to begin with the **<surface>** tag and end with the **</surface>** tag. If you are starting a transcription from scratch and want to delete the OCR produced text, make sure that you delete everything that appears **inside** the **<surface>**

</surface> tags. The 'extra content at the end of the document' error means the surface tags have been deleted.
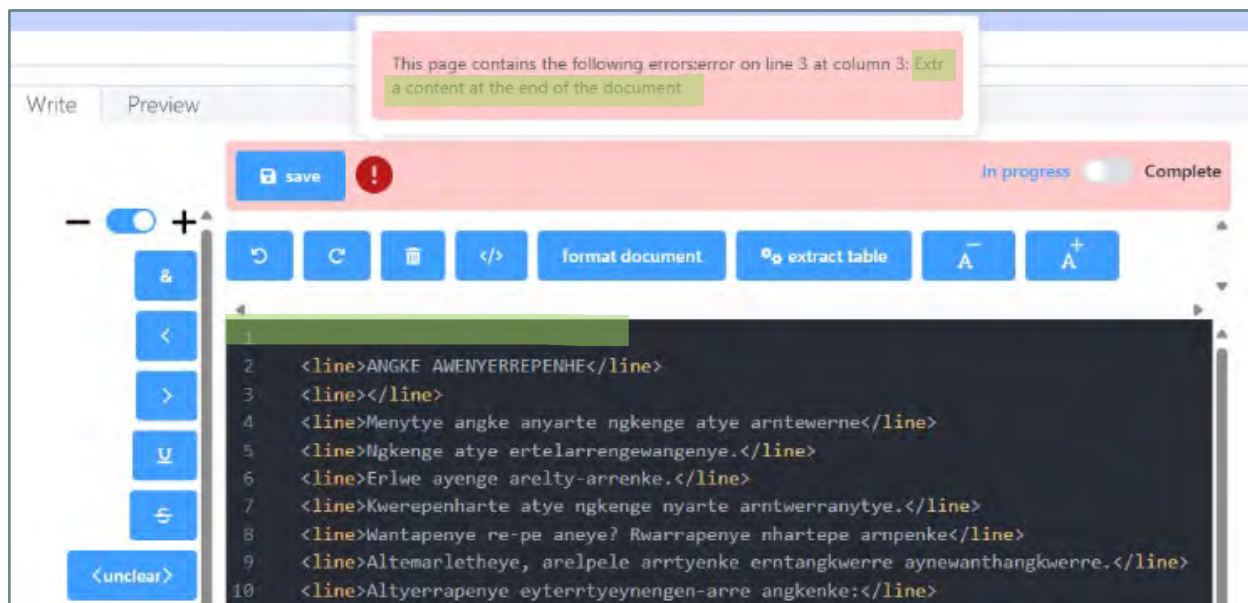
If you can, click the undo button to undo the deletion of surface tags. If this doesn't work, copy and paste the surface code from another page of the document, replacing the image file name with the page you are working on.

'xmlParseEntityRef: no name'

Another common error cause is using ampersand (&) in the text. This cannot be done using your keyboard as the ampersand is used within the coding language.
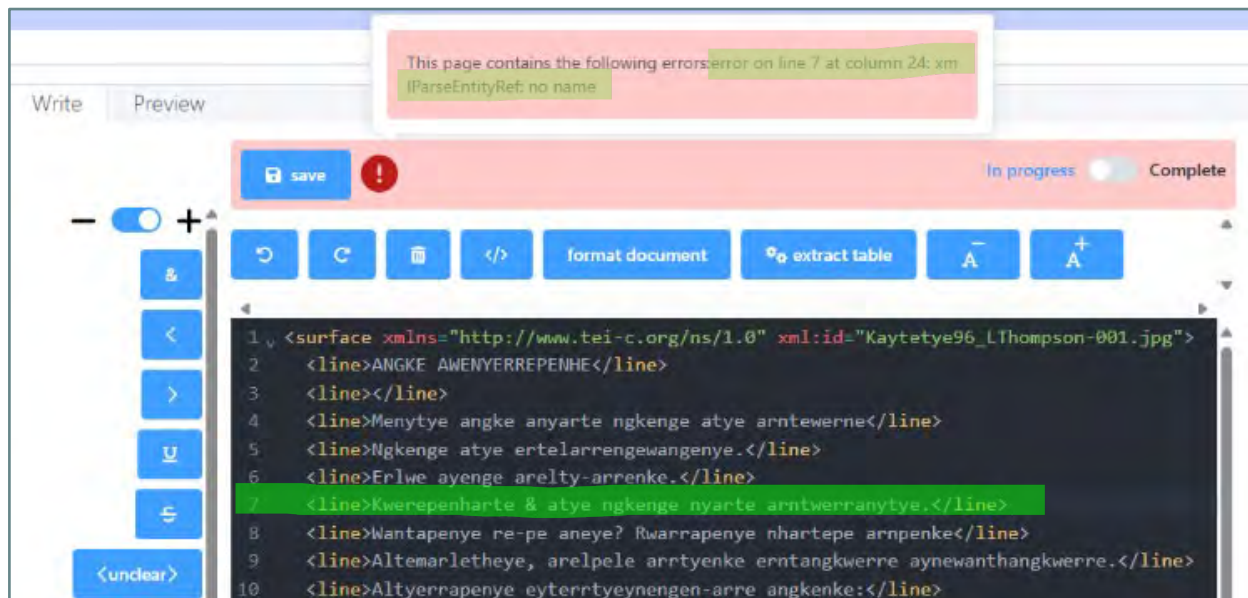
To insert an ampersand, use the blue '&' button. Once the ampersand is inserted with the mark up button, you will be able to save the transcription as normal.
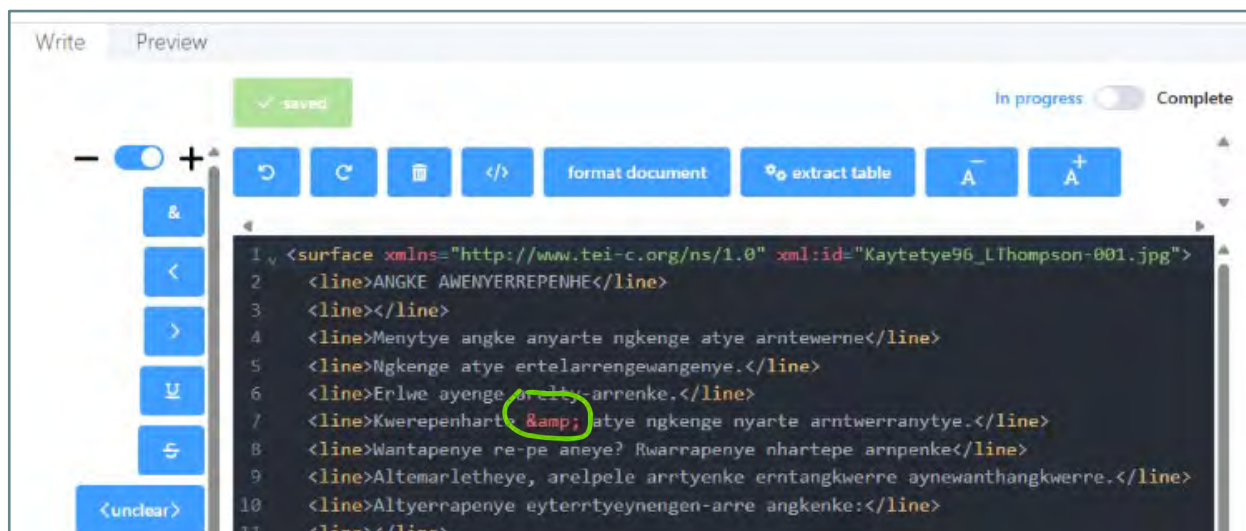
'Invalid element name'

Like ampersand, chevrons (< and >) are also used in the TEI coding language. If you type a chevron into the text, you will receive the TEI error message 'invalid element name'. Note the yellow and red colour coding in the screenshot below as the Workspace interprets words that follow the chevron as part of the code.
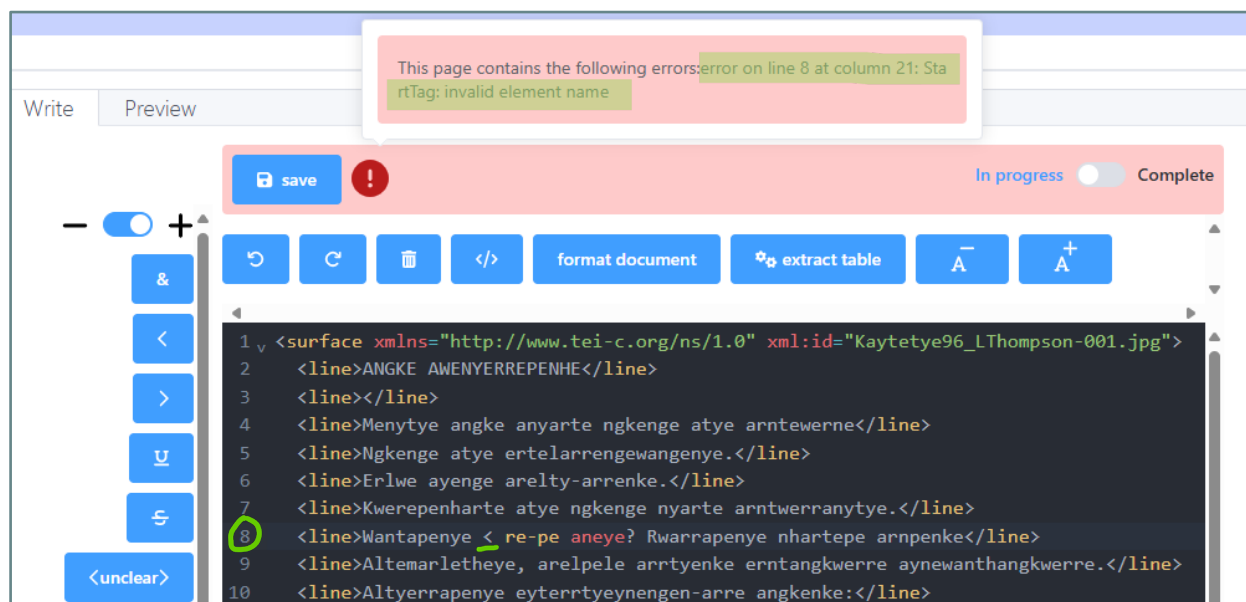
To insert a chevron that appears in the text, use the blue '<' and '>' buttons in the sidebar.

4.6 Advanced TEI

In the Nyingarn Workspace, TEI XML markup is added through the blue buttons on the lefthand side of the transcription editing box. In addition to these functions, some XML commands can be added manually. Some of these advanced XML commands are listed below:

<milestone> element

The <milestone> element can be used to create columns. Insert a <milestone> element at the point in the text where the column layout changes. The element should have a unit attribute whose value should be "column", and an n attribute whose value should be the number of the column, counting from left to right. For two-column texts, you will therefore have elements like <milestone unit="column" n="1"/> and <milestone unit="column" n="2"/> in alternation.

NOTE: <milestone> elements can end with a /> so opening and closing your column with <milestone></milestone> is not necessary.

<space> element

For significant white space (such as an indent), use a <space> element with a quantity attribute, e.g. <space quantity="5"/> to insert 5 character widths of space.

<term> element

The <term> element can be used to tag a language in the manuscript. Insert the following around the part of text you would like to markup: <term xml:lang=" "> </term>. The language name goes between the " ".
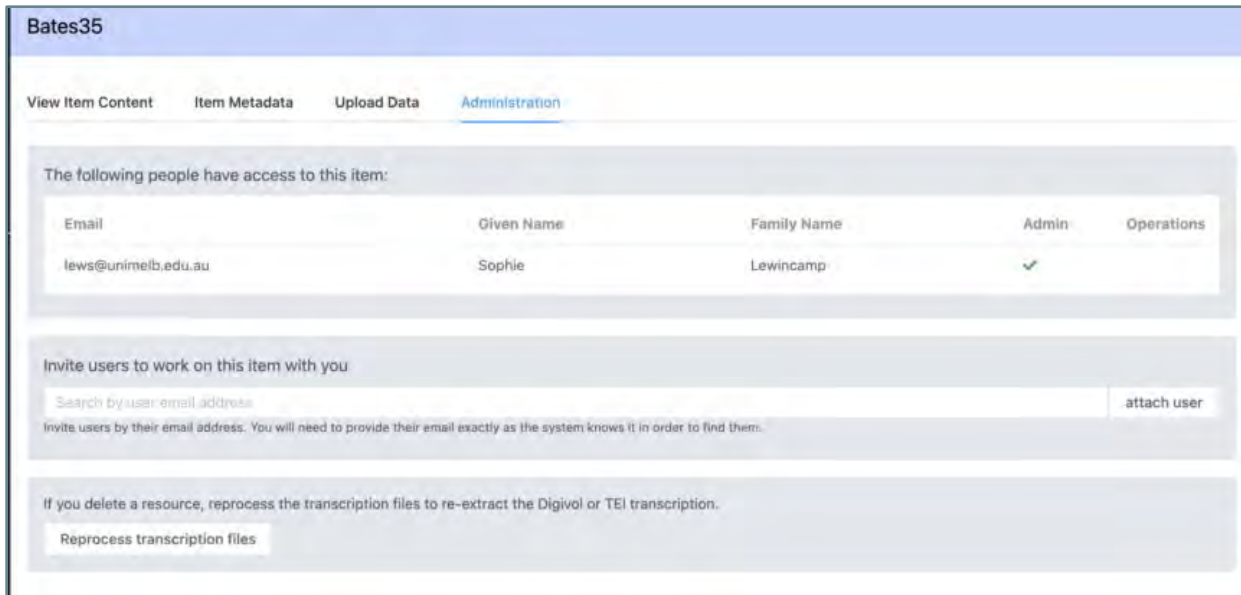
<gap> element

If there is a portion of the manuscript that is unreadable, this can be tagged as <gap> </gap>. In the PREVIEW tab this will present as '[illegible]'. You can use this element where you have no idea what the manuscript says, compared to the <unclear> </unclear> tag which indicates you can at least guess at what was written in the original manuscript.

4.7 Sharing your work

If appropriate, you can share your item(s) with other Nyingarn Workspace users. Click the *Administration* tab once your item is created. Then type the user's email address into the 'Invite users' box and click attach user (right hand side).

Why can't I add a user to my manuscript?

If you are typing in an email address and trying to 'attach user' to your manuscript, but nothing is happening make sure that:

1) You have typed the email address correctly.
2) The user has been added as a permitted user by a Nyingarn admin.

5. Metadata

Once you have created an item in the Workspace, you can add metadata.

Click on the *Item Metadata* tab. The Nyingarn Workspace has a standard metadata schema. The schema profile displays the fields our team has developed in consultation with AIATSIS. There are four sections of metadata – About, Original Source Information, Who and Other.

The *About* tab contains the core data for your manuscript item; see below.
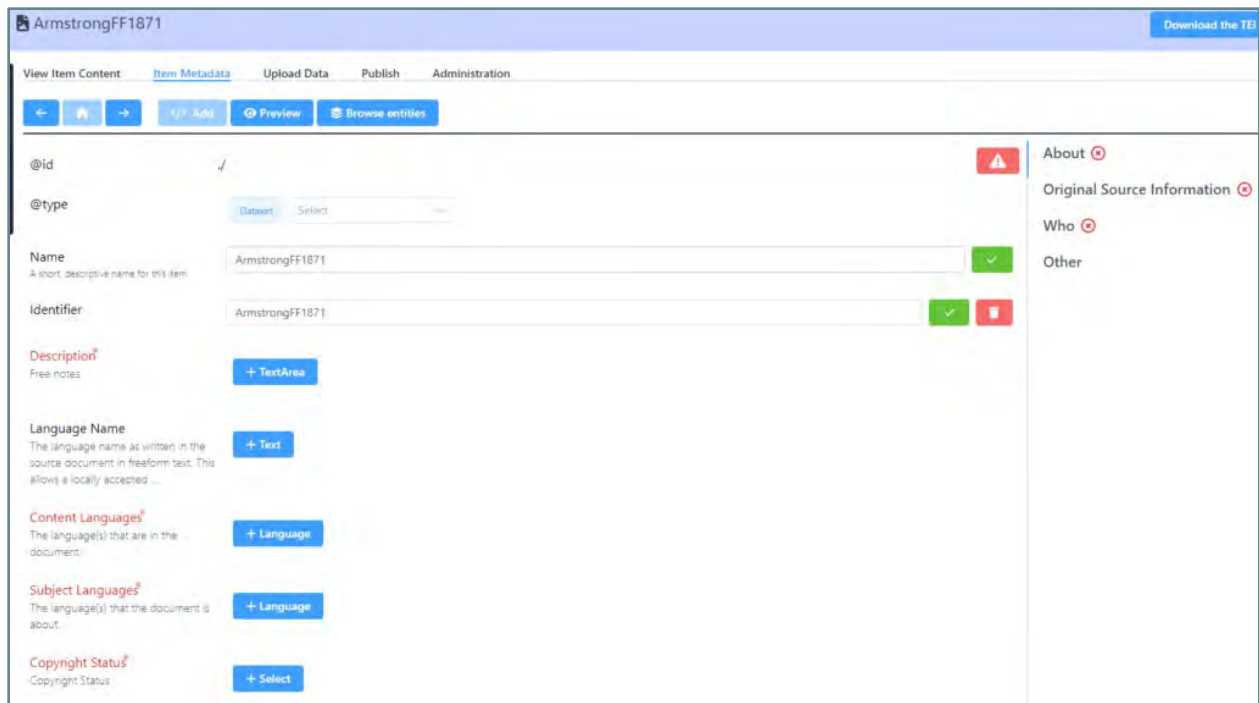
The location of the manuscript language is recorded and displayed on a map, in the *Other* tab.
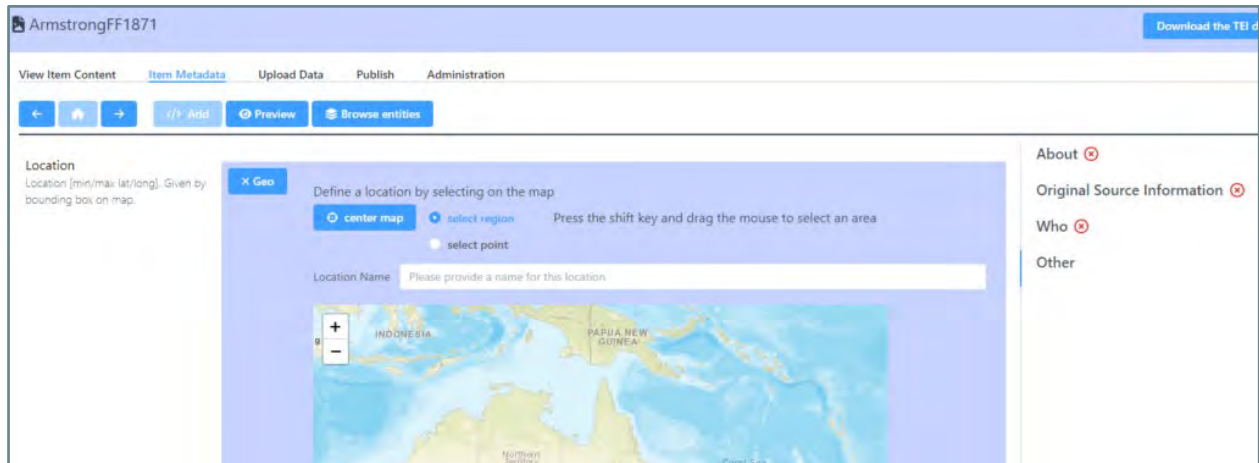
The *Original Source Information* tab connects your manuscript with existing catalogues and publications.
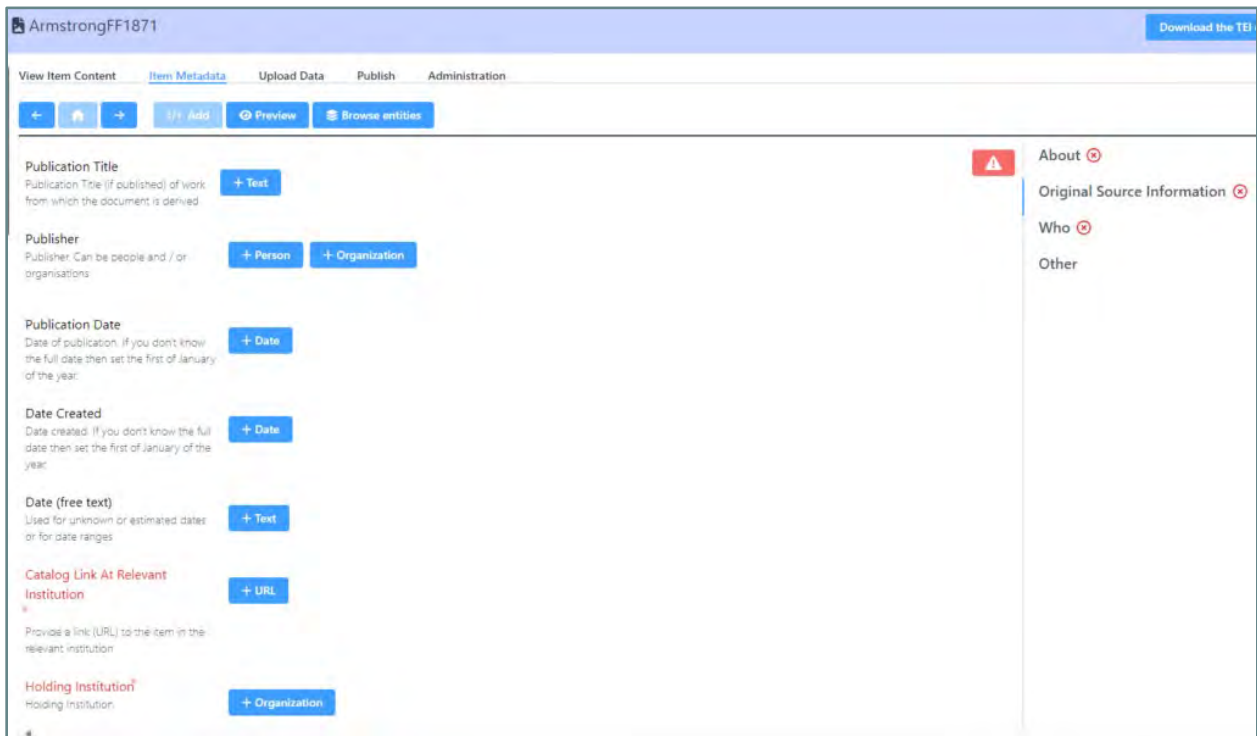
The *Who* tab details who worked on the manuscript. Metadata fields in red guide the minimum metadata needed for an item to move into the Nyingarn Repository.

6. Gaining and uploading permissions

Before an item can move from the Workspace into the Repository, two permission documents need to be signed and uploaded. Permission documents can be found on the nyingarn.net site.

1. Language authority permission
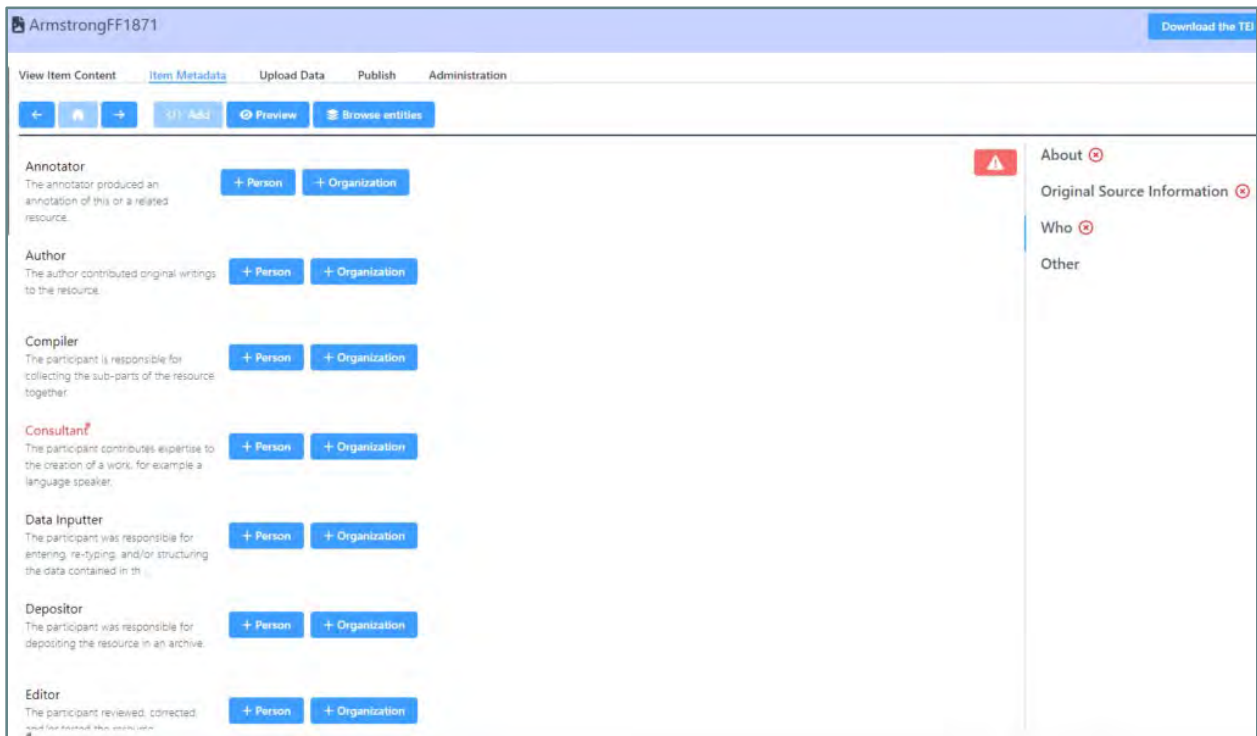2. Copyright holder or equivalent

Terri Janke Company has designed these documents in consultation with the Nyingarn Team. These forms are made available under a Creative Commons Attribution ShareAlike licence.

If you are ready for your item to move into the Repository, you can upload your permission documents in the 'Publish' tab.

To upload your files, click *Choose files* for each permission document. The file must be named according to the naming convention. For the example document above, the copyright holder file would be named *Bates36-rights-holder-permission.pdf* and the language authority permission file would be named *Bates36-language-authority-permission.pdf*.

7. Extracting your transcription work

7.1 Downloading your transcription as a TEI file

Once you have completed your transcription, it can be downloaded from the Nyingarn Workspace as a TEI document. Navigate to your item and click 'Download the TEI document' in the top right corner. The document will appear automatically in your downloads folder.
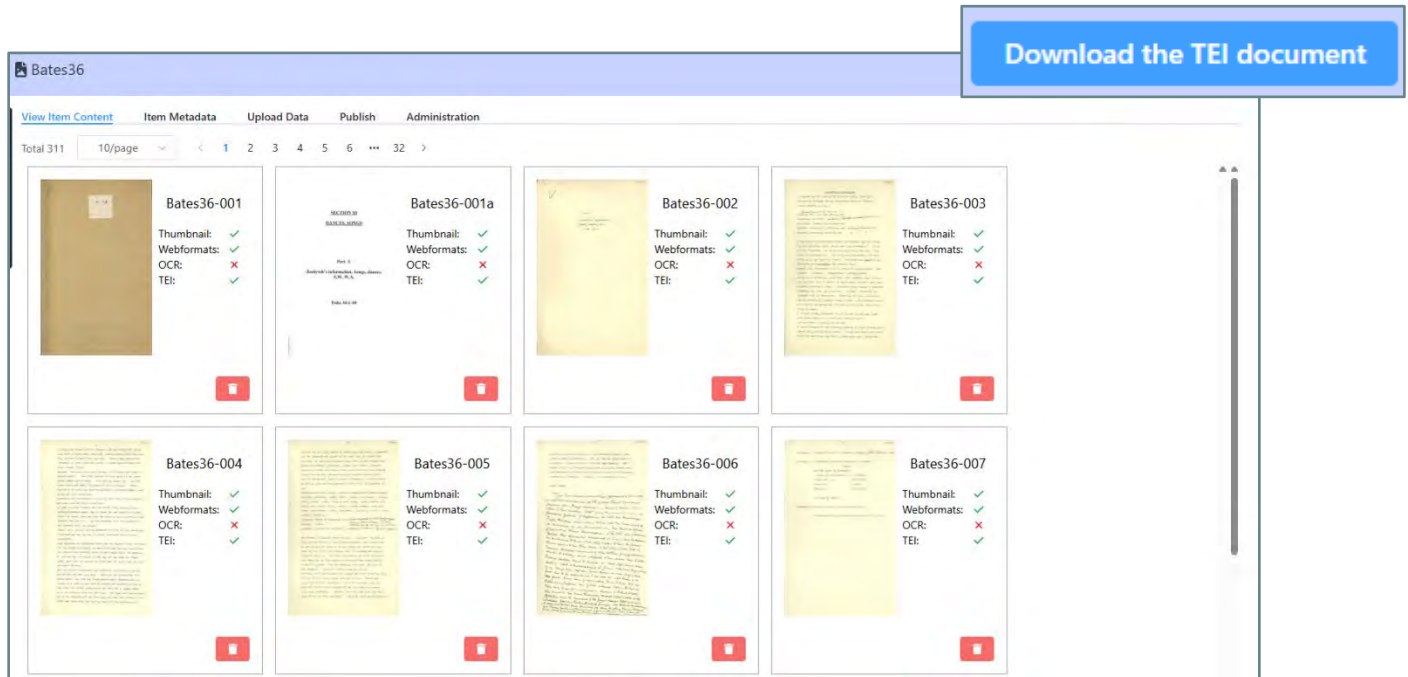
7.2 Converting your TEI transcription file to a PDF

To convert your downloaded transcription into a PDF or Word document, use an online TEI conversion tool such as TEIGarage (https://teigarage.tei-c.org). For TEIGarage, follow the steps below.
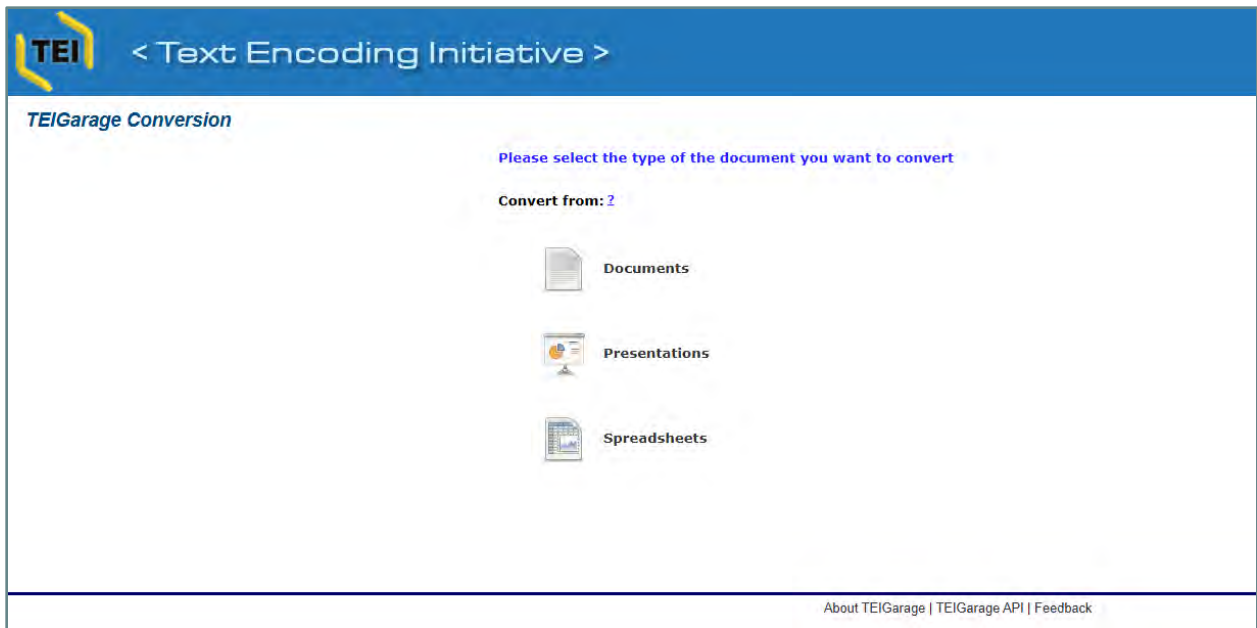
Step 1. Open TEIGarage in your browser. Select convert from *Documents.*
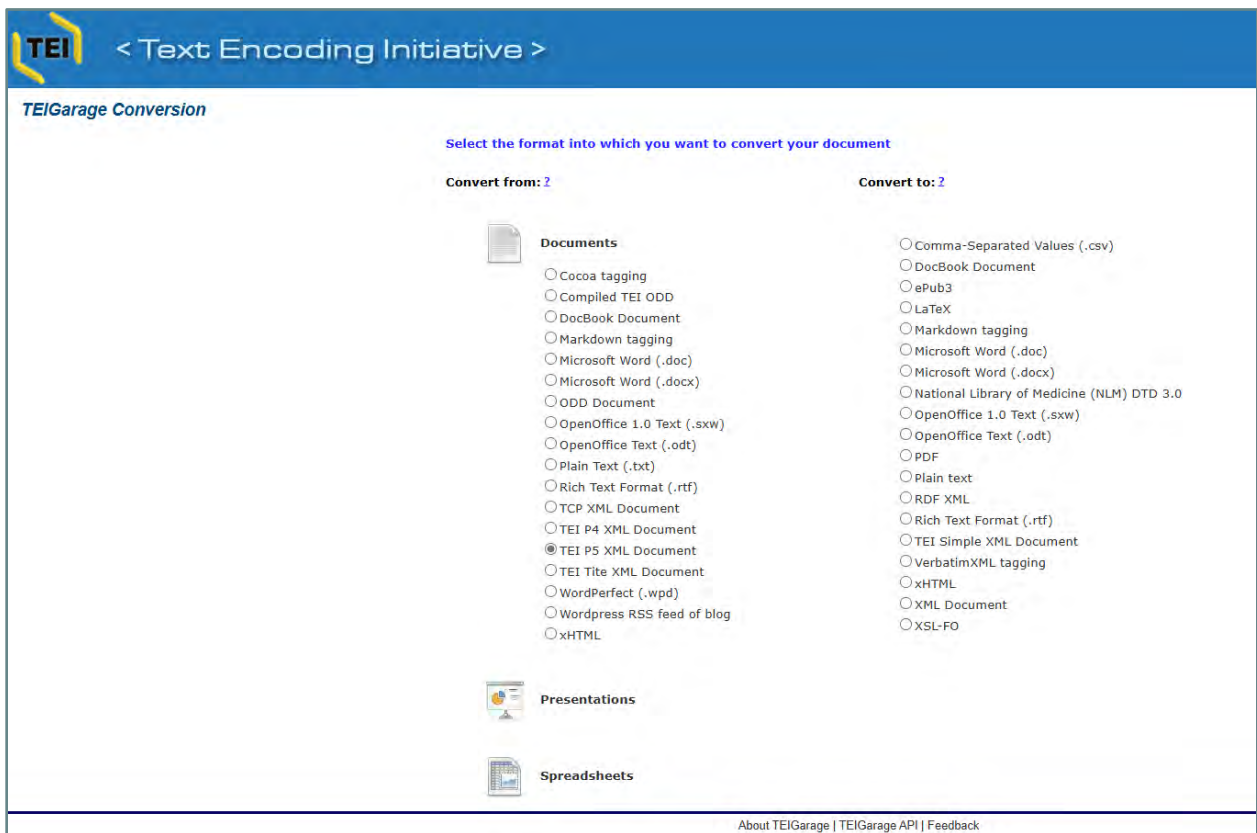
Step 2. Choose to convert from *TEI P5 XML Document,* converting to your preferred format (PDF or Microsoft Word .docx).

Step 3. Once you have selected your formats, upload the TEI file to the *Select file to convert* section and click *Convert*. Your converted file (.pdf or .docx) will appear in your downloads folder as soon as it is ready.

7.3 Bulk editing the TEI document

If you want to, you can edit the XML document. This can be useful if you want to make global changes, especially in long documents, instead of editing each individual page in the Workspace.
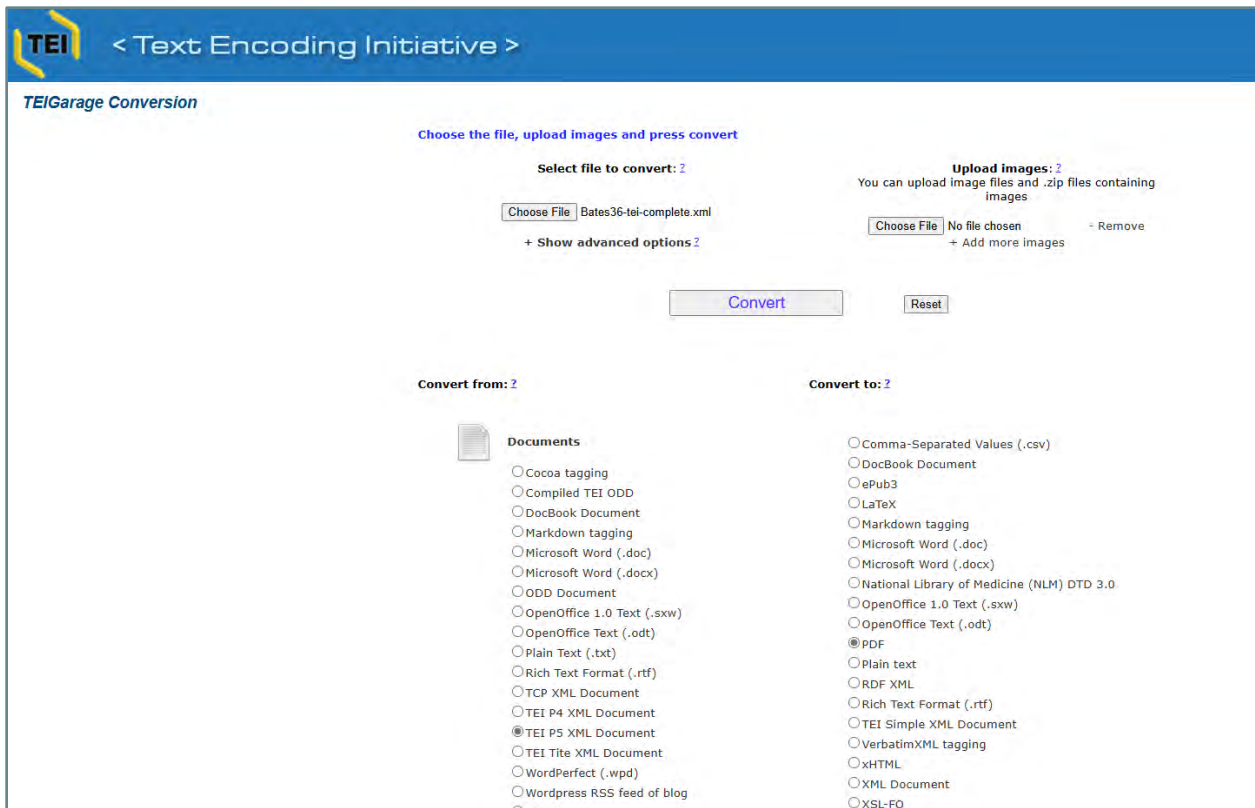
For example, you may want to code the language words and the English words. The following text comes out of a transcript: *<p> ládá – scudo shield </p>*. In this example there are three languages: Ngarla, Italian, and English. To code each of these words in the whole document, we can use regular expressions rather than going through each line and manually entering codes. The result will be as follows:

<p><term xml:lang="x-austlang-W40">ládá **</term>** – **<term xml:lang="it">**scudo **</term><term xml:lang="en"><supplied>**shield**</supplied></term></p>**

Notice that we have added **<supplied>** for information that is not in the original document, but is useful, in this case, an English translation.

Once you have made all the edits you want in the XML document you can reimport it to Nyingarn. This time you need to select the button to overwrite any existing transcriptions. This will delete the existing transcription and replace it with the one you have edited outside of Nyingarn.

8. The Nyingarn Repository

The Nyingarn Repository is a space where you can share your manuscript transcriptions more widely, with access restricted to a set of users selected by you or open access. For items to go from the Workspace to the Repository, they must have accurate metadata, community (or communities) permission, and an accurate transcription.

8.1 Sending items to the repository

Once you are satisfied with your manuscript transcription, have attained and uploaded community permission(s) and entered the metadata for your item, navigate to the 'publish' tab in your item.

You will see that the Workspace has accepted your permissions documents. Click on 'Generate the complete TEI document' to prepare your manuscript transcription.

Permissions   Permission forms loaded

TEI Document   **Generate the complete TEI document**

Visibility   **This item is open access**   ⬤   **This item is restricted to a defined set of users**

In both cases, users will need to agree to the Nyingarn Terms and Conditions of access.

Access Narrative

Detail the reasoning for restricting access to this item.

Review Date   📅 Pick a day   After this date the item will become open access.

Authorised Users

List the email addresses of the people who are allowed to have access to this item. One per line and/or separated by comma.

Publish this item   Cancel

Specify whether you want your item to be open or restricted access using the Visibility toggle. If you choose restricted access, you will need to enter an access narrative detailing reasons for restricting access, and list the authorised users by email address. You can also elect a review date after which the item will become open access (optional).

Once you have chosen the access type, click 'publish this item'.

The Nyingarn administrators will receive a notification that an item is awaiting review before moving to the Repository. They will make sure that transcription is complete and that your item has appropriate metadata and permissions, and then will confirm that the item can be published to the repository.

8.2 Using the repository

The Nyingarn repository can be accessed here: https://repository.nyingarn.net/. Use the map view to search items in the repository by area. You can also filter by language and search within manuscripts.

Open access

Manuscripts that are open access can be accessed by anyone using the repository. They do not require a login to view. The person viewing an open access item in the repository can see the original manuscript pages, the transcription, and the metadata.

Closed access

Manuscripts with restricted access require users to be logged in to the repository and linked to the manuscript for them to be able to view it. To log in to the repository, click 'login' in the top right corner. Enter your email address. A login code will be sent to your inbox.
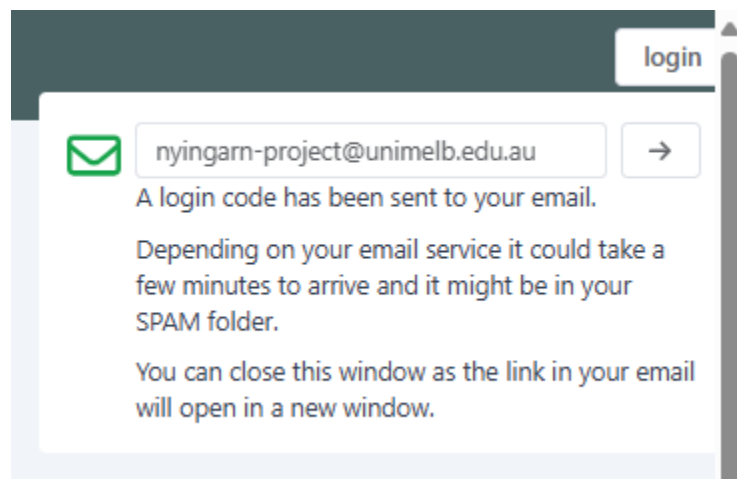
Users who have not been given access to a restricted manuscript will still be able to see the title and basic metadata for that item.