

Edgeworth TEI Workflow

Developers: Eliza Wilcox, Meredith Hale, Hilary Havens, Jamie Kramer

Version: 3.0

Date: November 2022

1. Before beginning work on the project as a whole, be sure you are a member/owner on the [Maria Edgeworth Github repository](#). If you have not been added, contact Meredith.
2. Before beginning work on any files, make sure your GitHub desktop client is up to date. If using GitHub in a browser or working with git on the command line, disregard this step.
3. Open the appropriate textfile folder for the letters you are working on. Textfiles can be located through Edgeworth → Data → Edgeworth_Data_Library/Archive → Library/Archive_textfiles
 - a. For example, for the Beinecke textfiles, you would go to Edgeworth → Data → Edgeworth_Data_Beinecke → Beinecke_textfiles
4. Open the [Taxonomy Terms](#) spreadsheet.
5. Open the [Edgeworth Letter Holdings in Archives Worldwide](#) spreadsheet.
6. Steps 2-4 will be the places to find the textfile for the letter you are encoding (step 2), the taxonomy you need to encode (step 3), and information about the archive necessary for the template (step 4).
7. Download the most recent [TEI template](#) from Github.
 - a. If there has not been an update since you last used the template, you can open it directly from a saved copy on your own computer.
 - b. If you are working on editing an already existing XML document, download it from the appropriate Drive or GitHub folder.
 - i. Avoid downloading and working on existing XML documents from Github if there are open pull requests. Doing so can create merge issues later on and erase work that's been done.
8. Select the letter you are working on from the Library/Archive_textfile.
9. Go to Zooniverse and [login to the MELP ALICE](#) page.
 - a. Once in Alice, click "Standard Volunteer Workflow."
 - b. Find the Zooniverse folder that corresponds to the Library/Archive_textfile that you're working with.
 - i. For example, if you are encoding "Edgeworth_Beinecke_330_1817_02_09_page01.txt" then you need open the folder "Beinecke 1" and find "Edgeworth_Beinecke_330_1817_02_09_page01.jpg" in Alice.
10. If working on a new letter, create a new file within your local repository, copy the most recent MELP TEI template into the document, and retile it using the naming convention detailed below.
 - a. General git hygiene should be followed when working with the Maria Edgeworth repository.
 - i. At the beginning of your work session, check the Git repository to ensure there are no pull requests open. If there are, work on files **on our own local repository** and do not pull them from Git. Pulling files while a pull

request is open can create merge issues later on and erase work that's been done.

- b. If there are no pull requests open, pull updated files from the main branch to ensure you are working with the most updated versions of each file.
 - c. When working on the files, keep track of changes made so those changes can be clearly described when committing the file to the main branch.
 - i. General git hygiene calls for committing when making each unit of change. On this project, we suggest committing the file as you complete each section of the template, and once the file is complete, making commits for any change, even incremental ones, so as to keep track of editorial changes.
 1. For example, when initially creating the file, commit the file, describe it as "Created Edgeworth_Beinecke_330_1817_02_09 but did not finish filling out the template," and then push it to main
 2. For another example, as you finish each section of the file, commit the file, describe it as "finished encoding the text of the letter and updated the taxonomy terms sheet to correspond to new terms," and then push it to main
 - a. The key and similarities between these examples is being **specific and descriptive** and what has been done to the file
11. Retitle the XML document to match the Zooniverse text file **but remove the page count from the XML document title**.
- a. For example, if working on a text file from Zooniverse labeled "Edgeworth_Beinecke_330_1817_02_09_page01" then the XML document should be titled "Edgeworth_Beinecke_330_1817_02_09.xml"
12. Add the primary XML editor's name to the "Editor" line formatted as firstName lastName.
- a. If this has already been added by another editor who is the primary editor, skip this step.
13. List all other transcribers in the responsibility statement [<respStmt>] of the title statement section [<titleStmt>], and separate each person's name with <persName ref="Id_who"></persName>
- a. For example, <persName ref="Id_who">Jane Smith</persName>
 - b. Whoever the *most recent person* to interact with the file is should add their name to this last list.
14. In the edition statement [<editionStmt>], enter the date the XML was first created and saved to the Google drive/GitHub.
15. In the responsibility statement [<respStmt>] of the edition statement section [<respStmt>], list who took the photos of the letters being encoded *and/or* who gave permission for use of the photos using the <orgName ref="Id_who"> tag.
- i. Consult the "Organizations tab" in the [Taxonomy Terms](#) spreadsheet to get the correct form of the organization name (and URI). The term will go between the <orgName> tags and is listed in the "Term" column on the "Organizations" sheet.

- ii. Copy and paste the whole URI to replace "Id_who". The URI is listed in the "URI" column on the "Organizations" sheet. An example URI is <http://id.loc.gov/authorities/names/n80020306>
16. Next, in the same section, make a comma separated list of each .jpg photo file name associated with your letter,
 - a. For example, Edgeworth_Harvard_01-01-1850_page01.jpg, Edgeworth_Harvard_01-01-1850_page02.jpg, Edgeworth_Harvard_01-01-1850_page03.jpg
 - i. This is the standard file naming convention for the *Maria Edgeworth Letters Project*, so if the JPGs you are working with *are not named this way*, they need to be renamed and resaved with those file conventions.
17. In the publication statement section [<publicationStmt>], enter the organization name used in step 13 here <!--Reproduced by courtesy of the <orgName ref="#OrgName">Name of Org</orgName>-->
 - a. For Name of Org, follow the same steps as 15i.
 - b. For #Orgname follow the same steps as 15ii.
 - c. Remove the <!-- --> brackets.
18. Change the <idno> tag in the publication statement to match the file name.
 - a. For example, if the file name is "Edgeworth_Beinecke_330_1817_02_09.xml" then the <idno> tag should read
<idno>Edgeworth_Beinecke_330_1817_01_24</idno>
19. For the manuscript source in the header, go to the Edgeworth Google Drive, locate the spreadsheet titled "[Letter Holdings in Archives Worldwide](#)," and find the letter being encoded.
 - a. I find the best way to do this is the type in the sender/receiver (whichever is not Maria Edgeworth) and then cross-referencing for the date.
20. Then, using "Archival Institution" column from the spreadsheet, add the original manuscript source to the manuscript description [<msDesc>] in the source description [<sourceDesc>]
 - a. For example, if the source is Yale Library, then you'd add Beinecke Rare Books Library to the manuscript description.
21. Then, using the "Ms Collection Name" column from the metadata spreadsheet, add in the manuscript collection to the <collection></collection> tag
 - a. For example. If the collection name is Abbotsford Collection, add Abbotsford Collection to the <collection></collection> tag
22. **Optional:** Add OCLC numbers from Worldcat into idno (or significant identifiers from libraries that hold the materials). You can find OCLC numbers by searching here - <https://www.worldcat.org/> - and copying the numbers after "oclc" in the url.
 - a. E.g. for <https://www.worldcat.org/title/horses/oclc/1158483014> the OCLC identifier is "1158483014"
23. **Optional:** add in a physical description of the letter to the <physDesc> tag.
24. **Skip for Now:** In the bibliography section, add the complete information for where the letter was previously published to the <bibl> tag.
 - a. Where should they go to find this information? – Decision tabled for now.

25. In the encoding description [<encodingDesc>], add the sender's name and the recipient's name in the appropriate areas. Each should be added firstName then lastName.
26. If you are revising previous edits, record the revisions made and the date they were made in the revision description [<revisionDesc>] using the <change> tag.
27. For encoding the transcription of the letter itself:
 - a. In the opener section [<opener>], edit the <dateline> tag for the letter. If there is no dateline, omit the element.
 - b. If the letter opens with a salutation, edit the <salute> element using the <persName> tag. If there is not salutation, omit and delete the element
 - i. We are using a personography for people's names, so tag names with <persName ref=../Personography.xml#ExamP1>
 - ii. IDs for people can be found in the Taxonomy Sheet
 - c. Transcribe the body of the letter immediately after closing the <opener> section.
 - i. First, separate the text into appropriate sections. You will need to consult the Zooniverse image files to identify which part of the transcription go in each section: <body>, <closer>, <postscript>, <address>
 - ii. Once separated into appropriate sections, add a self-closing page break element [<pb/>] for each photo file, and use n and facs functions to describe each letter.
 1. For example, for a one page letter from Maria Edgeworth to Richard Bentley in 1848, there should be a <pb/> for the singular letter: <pb n="1" facs="Edgeworth_NLS_3456_1848_04_10_page01.jpg/>
 - a. The "n" changes to correspond with the page number of the letter. Page 1 in the image files could be the address sheet, which would mean "<pb n="1" facs="Edgeworth_NLS_3456_1848_04_10_page01.jpg/>" would be added to the <address> section and NOT the body.
 - b. When there is a page break, record it WITHIN the body of the appropriate paragraph using the self-closing page-break element.
 - i. Example: "Seeing you recently was enjoyable, and I wish we <pb n="2" facs="Edgeworth_Beinecke_330_1817_01_24_page01"/> could do it again soon."
 - ii. In this example, the self-closing page break indicates that the sentence continues on page 2 (n="2"). If it were page 3, n would be n="3".
- iii. Once all page breaks have been added, use the Zooniverse image files to figure out where to add paragraph breaks.
- iv. Once all paragraph breaks have been added, begin encoding features of the letter as well as tagging for person, place, and works.

1. Some common XML elements used to encode features of the letter
 - a. Insertions <add>, deletions , unclear <unclear>
 - b. Superscript: <hi rend = "superscript"> ... </hi>
 - c. Underline: <hi rend = "underline">...</hi>
 - d. Change ampersands to &
 - e. For abbreviations **and** when words go across line breaks, the abbreviated or broken text should be regularized within <choice><orig>**broken text or abbreviation**</orig><reg>**combined text**</reg></choice> so that a full text search will return the letter, with the original text enclosed by an orig element
 - i. Example: The enclosed is a
 <choice><orig>mat</orig>
 ter</orig><reg>matter</reg></choice> of business
 - f. Special cases
 - i. Figures and drawings: <figure> at moment they appear with <figDesc> inside containing a short, prose description of the item. See <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-figure.html> for more examples.
 - ii. Tables: <table> at moment they appear with optional <row> and <col> separators. See <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-table.html> for more examples.
2. MELP uses an in-house taxonomy to tag persons, places, and works. These “ographies” – the personography, placeography, and workography – are living documents hosted via the Taxonomy Terms sheet. Below are details on how to tag persons, places, or works using the “ographies” as well as instructions on how to add to each “ography” should an encoder come across a new term not previously defined in the “ographies.”
 - a. <persName> procedures
 - i. Names are tagged using a <persName> and a @ref tag:
 1. <persName
 ref=../Personography.xml#ExamP1>
 - ii. IDs for people can be found in the Taxonomy Sheet.
 - iii. If you come across a new name that needs to be added to the Personography, please follow these steps:
 1. In Column A, add the person’s name in LastName, FirstName convention.

2. If possible, look up and add the person's birth date to the column next to the name.
 3. In Column B, add the ID tag using the naming convention.
 - a. The naming convention is the first four letters of the last name plus the first letter of the first name and then a number to create a unique six character ID.
 - b. For example, Mr. John Smith and Mrs. Jane Smith would be tagged SmitJ1 and SmitJ2, respectively. Three letter names are given in full and will yield five character IDs rather than six.
 - i. For example, Mr. John Doe and Mrs. Jane Doe would be tagged DoeJ1 and DoeJ2
 4. Go to <https://viaf.org> and locate the VIAF URI for the person, and add the VIAF URI to column C.
 - a. If you cannot locate the person on VIAF, please note this in the "Notes" column as well as noting the letter this person appears in. Once your note is written, be sure to tag an editor so that we can try and find more information to provide an appropriate URI.
 5. In column D, put your initials to indicate you were the person to add this individual to the personography.
- b. <placeName> procedures
- i. Names are tagged using the <placeName> and a @ref tag
 1. <placeName
ref=".Placeography.xml#placeName">place
Name</placeName>
 2. For example, Edgeworthstown would be
<placeName ref=
"./Placeography.xml#Edgeworthstown_IE">
Edgeworthstown</placeName>
 - ii. IDs for places can be found in the Taxonomy Sheet.

- iii. If you come across a new place that needs to be added to the Placeography, please follow these steps:
1. In Column A, add the place using the language from the text.
 - a. For example, if it says “Bath,” add “Bath” not “Bath, England.”
 - b. If the location is a country, place it in column A and leave column B blank.
 2. In Column B, add the location’s country.
 - a. For example, if Column A is “Dublin,” add “Ireland” to column B.
 3. In Column C, add the ID tag using the naming convention.
 - a. The naming convention is name of the specific place conjoined with the country code using en dashes to separate words and an underscore to separate the place from the country
 - i. For example, London would be given the ID London_ENK
 - ii. Country IDs are defined by geonames and can be located there
 4. Then, go to www.geonames.org and use the search function to try and locate the place
 - a. If you find the specific place, copy and paste the longitude and latitude coordinates into Column D
 - b. Then, download the URI by clicking “rdf” and downloading the URI.



- c. This should open in oXygen or other software capable of reading XML,

and you can copy and paste the URI from there into Column E.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<rdf:RDF xmlns:cc="http://creativecommons.org/ns#" xmlns:dcterms
  <gn:Feature rdf:about="https://sws.geonames.org/2643743/">
```

5. If there are notes, please add those in the “Notes” column in Column G
 6. Add the initials for the editor who added the term to Column H
 7. Add the letter where the term was located when it was first added to the sheet in Column I
- c. <title> procedures
- i. Works are tagged using a <title> element and a @ref tag:
 1. <title ref=”./Workography.xml#title1”>Title</title>
 2. For example, *Belinda* would be <title ref=”./Workography.xml#Belin1>Belinda</title>
 - ii. IDs for works can be found in the Taxonomy Sheet
 - iii. If you come across a new work that needs to be added to the Workography, please follow these steps
 1. In Column A, add the title of the work
 2. In Column B, add the foreign or alternative title that also refers to the same work if there is any foreign or alternative title
 3. Using www.wikidata.org, locate the title and add the URI from wikidata to Column C
 - a. If the work cannot be located on wikidata but the author is known, add the author’s name to column E before proceeding to column K.
 - i. In Column K, add any notes about the work and tag an editor so the work can, hopefully, be identified. Indicate the specific letter where the work has been referenced.
 4. If possible, find a link to the full text and insert the link into column D.

5. In Column E, add the author's name, first and last, and their birth and death date, if known
 - a. This information may be located in the Personography, so be sure to cross reference
6. In Column F, add the author's VIAF URI
 - a. This information may be located in the Personography, so be sure to cross reference. If not in the Personography, follow the instructions at 4a under "<persName> procedures"
7. If there is a known collaborator, add their name(s) to Column G, using the same format as the author's name and birth/death date as detailed in step 5 above
8. If there is a known collaborator, add the collaborator's VIAF URI, using the same procedures and format as detailed in step 6 above
9. In Column J, add the work's ID
 - a. The ID naming convention for works is to take the first five letters of the work and join it with a number, beginning with 1 and working up sequentially as necessary and ignoring articles (a, the, an)
 - i. For example, *Belinda* would be Belin1. If there was a reference to a work titled *Belinger*, it would be Belin2. If there was a reference to *The Murders at Kinsington Palace*, the ID would be Murde1.
 - ii. If a work's first five characters are two separate words, the same conventions remain. For example, *The City of Ladies* would be Cityo1.
3. Retain all line breaks using the <lb/> element.
 - a. I typically wait to add the <lb/> element last.

- d. In the closer, add the closing portion of the letter and use all the same encoding steps as listed above.
 - i. For example, if encoding
Yours truly,
Maria Edge
 - e. Then your encoding should like
Yours truly,<lb/>
<persName ref=/.Persongraphy.xml#EdgeM1>Maria
Edgeworth</persName><lb/>
 - f. If you run into any uncertain places, names, persons, or dates, use the following to indicate the uncertainty by describing where the information comes from.
 - i. <persName ref=/.Personography#EdgeM1>Maria
Edgeworth</persName> went to <placeName
ref=/.Edgeworthstown_IE>Edgeworthstown</placeName>. She had
always liked <placeName
ref=/.Edgeworthstown_IE>Town</placeName>.
<note type="certainty">It is not
clear here whether or not "Town" refers to Edgeworthstown but context
leads us to be reasonably certain this is the appropriate
placeName</note>
28. Any postscripts should go in the <postscript> element *after* the closer and inside the <postscript></postscript> tag and follow all the same encoding steps as listed above.
29. Address information from anywhere on the letter goes *after* the closer in the <address> element. Use the self-closing tag <addrLine/> to separate addresses and use the same encoding steps as listed above.
30. Any other hands, including a librarian's or other persons, should be indicated in prose in a <note place="bottom">...</note>. Librarian/archive notes should not be in the main text. Eventually, we should talk about having a <handDesc> section in the header and possible IDs for different hands:
<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-handDesc.html>.