Generating Ography Files

1. Go to the TaxonomyTerms spreadsheet in the Edgeworth Drive
2. Assess the completeness of the "ography" file you are going to generate
   a. Does every row have a xml:id?
3. Download the TaxonomyTerms spreadsheet and delete all of the tabs except for the one you plan to create an "ography" file for. These instructions will create a personography file using the "People" tab of the spreadsheet.
4. Delete the second row of the spreadsheet (which contains unnecessary instructional information).

	A	C	D	E
1	Term	VIAF URI	Editor	Notes
2	(Lname, Firstname Mname, Birth date - Death date)		(initials)	(share any info/links to help identify individual)
3	Addison, Joseph, 1672-1719	http://viaf.org/viaf/7413288	JK	
4	Aikin, John, 1747-1822	http://viaf.org/viaf/59180443	HH	https://en.wikipedia.org/wiki/John_Aikin, brother of
5	Aikin, Lucy, 1781-1864	http://viaf.org/viaf/17362522	HH	https://en.wikipedia.org/wiki/Lucy_Aikin
6	Amici, Mr.		JK	https://en.wikipedia.org/wiki/Giovanni_Battista_An
7	Arago, François, 1786-1853	http://viaf.org/viaf/64023378	JK	astronomer, France https://en.wikipedia.org/wiki/F
8	Ashburner, Dr.			
9	B, Captain		JK	this is likely a person mentioned elsewhere, but "E
10	B, William		JK	
11	Babbage, Charles, 1791-1871	http://viaf.org/viaf/4963723		

becomes

	A	B	C	D	E
1	Term	ID	VIAF URI	Editor	Notes
2	Addison, Joseph, 1672-1719	AddiJ1	http://viaf.org/viaf/7413288	JK	
3	Aikin, John, 1747-1822	AikiJ1	http://viaf.org/viaf/59180443	HH	https://en.wikipedia.org/wiki/John_Aikin, brother of
4	Aikin, Lucy, 1781-1864	AikiL1	http://viaf.org/viaf/17362522	HH	https://en.wikipedia.org/wiki/Lucy_Aikin
5	Amici, Mr.	AmicX1		JK	https://en.wikipedia.org/wiki/Giovanni_Battista_An
6	Arago, François, 1786-1853	AragF1	http://viaf.org/viaf/64023378	JK	astronomer, France https://en.wikipedia.org/wiki/F
7	Ashburner, Dr.	AshbX1			
8	B, Captain	BxxxX1		JK	this is likely a person mentioned elsewhere, but "E
9	B, William	BxxxW		JK	
10	Babbage, Charles, 1791-1871	BabbC1	http://viaf.org/viaf/4963723		

5. Upload edited spreadsheet to OpenRefine by going to "Create Project" and then "Choose Files." Once the spreadsheet is uploaded, click "Next."

6. On the next screen, ensure that the data looks correct. Rename the project if desired and be sure to uncheck "Store blank rows" and "Store blank rows as nulls" in the lower-right-hand corner. Click "Create Project" in the upper-right-hand corner when satisfied.

7. Once the project is created, check the "ID" column for duplicate and blank values.
   a. To do this you can complete a "Text facet." Click the drop-down arrow on the "ID" column and then select "Facet" > "Text facet."

b. Then on the side facet, click "count" (rather than "name") and see if any ID values appear twice or more. If so, the duplicate ids will need to be distinguished from one another on both the OpenRefine project and back on the TaxonomyTerms spreadsheet. You will get an error in Oxygen later if duplicates are not resolved. The screenshot below shows a facet that requires no changes (if there are any IDs that appear twice, the "count" sort by would make them appear at the top).

c. You need to also check the facet to see if there are any blank values (or names without established xml:id values). Scroll to the bottom of the facet to see if any appear. If present, click on "blank" to see what names need IDs. Then create an ID based on the ID Conventions doc and fill in this value in the OpenRefine project and back on the TaxonomyTerms spreadsheet.

8. Next clean up the "Term" column values by trimming off any whitespace. To do this, click the drop-down arrow on the "Term" column, then select "Edit cells", "Commons transformations", and finally "Trim leading and trailing whitespace." This ensures that there are no extra spaces within <persName> in the resulting XML. Repeat this action on the "VIAF URI" column.

9. Next, use "Templating" and existing export templates on GitHub to create the XML.
   a. First find the associated export template on the Maria Edgeworth me-tei GitHub repository by going to the People_Places_Works folder. Each export template is named by the type of "ography" it is and then "open_refine_template." For instance, the file you need for people is named "personography_open_refine_template." Either clone the repository code to GitHub desktop or download the single export file you need.
   b. Open the file in Oxygen or MacDown. You'll notice that the template is composed of three separate sections. Those sections need to be copied and pasted into OpenRefine one by one.
   c. In OpenRefine, go to "Export" in the upper-right-hand corner and then click on "Templating…"
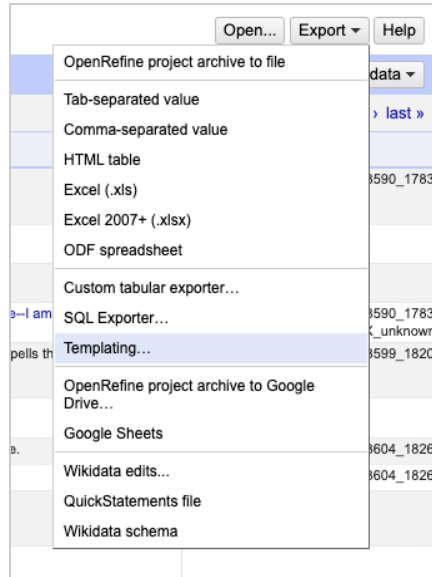
d. Next, fill out the "Templating Export" menu that pops up with the GREL code already present in the open refine export template you've opened from GitHub. Copy all the text (minus the back ticks ```) in ####Prefix to the "Prefix" section in the OpenRefine templating interface. Everything in the ####Body section goes to "Row Template". Delete the comma in the "Row Separator" box and leave it blank. Click "Export."

10. The exported file will be in your Downloads as a text file. Find and move the downloaded file to your Desktop and then change the extension (.txt) to .xml. Then open the file in Oxygen.
11. In Oxygen, ensure that you get a green box in the upper-right-hand corner and the message "Validation successful" at the bottom of the interface. If not, read and troubleshoot error messages. Save the file to your local GitHub repository and replace any older version of the "ography" you are working on with this new version (named the same thing as the previous file to track changes).
12. Transform the XML file using XSLT to streamline the formatting.
    a. Open the FormatTEI.xpr file in Oxygen from the People_Places_Works folder within the GitHub repository to achieve this. Then right click on the appropriate "ography" file that you just saved/created and go to "Transform" > "Transform with…"
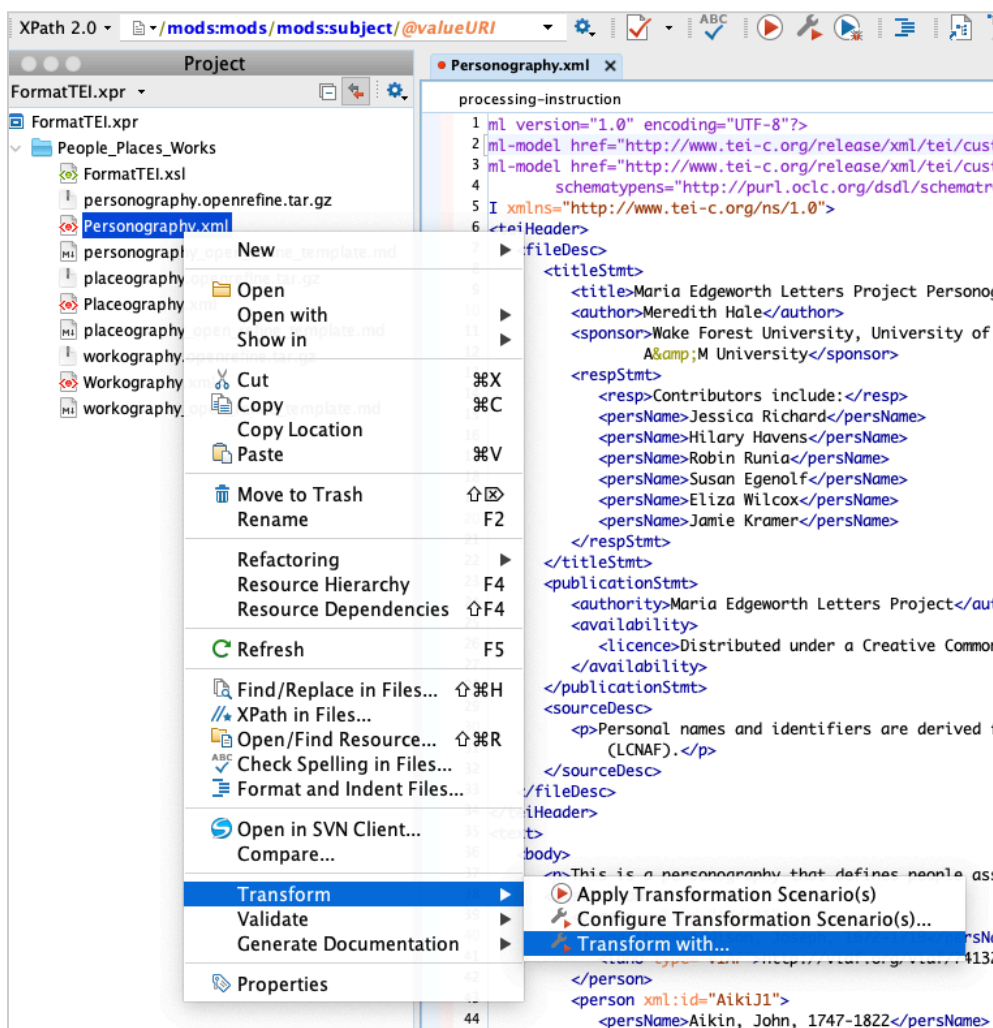
b. In the resulting pop up menu, select "" and then click "Apply selected scenarios".
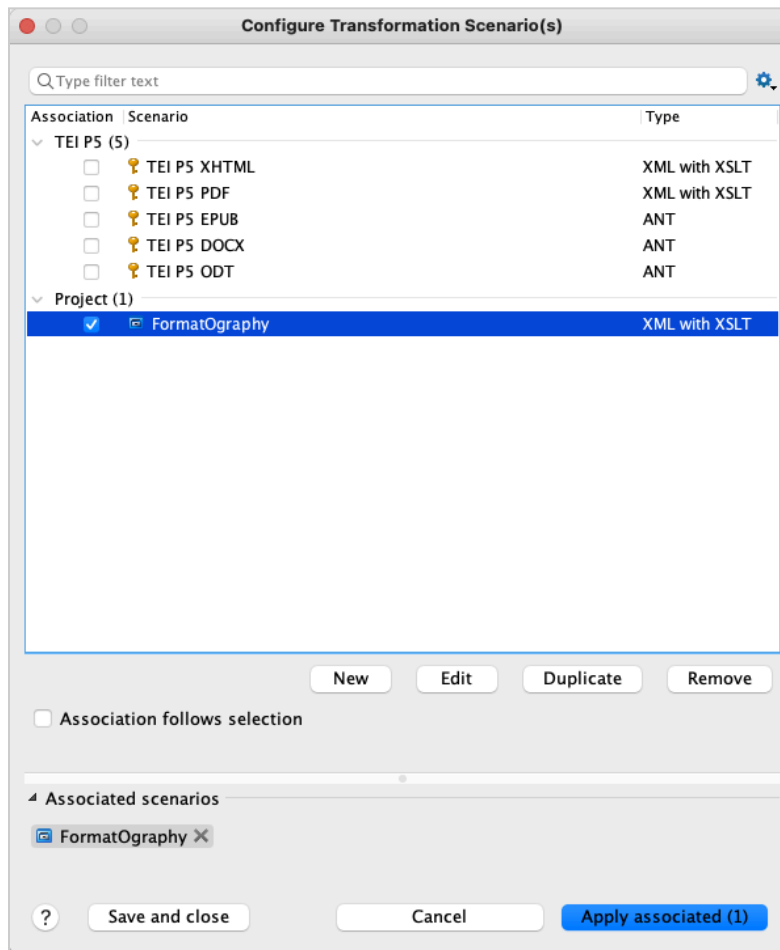
13. Save the resulting file to your local GitHub repository for Edgeworth and then commit the new "ography" file to GitHub.