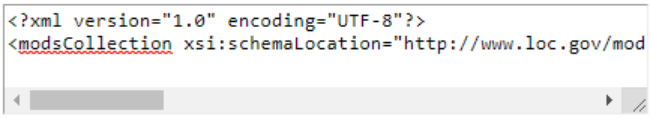
Metadata Transformation from WoS to Niner Commons Records

Overview

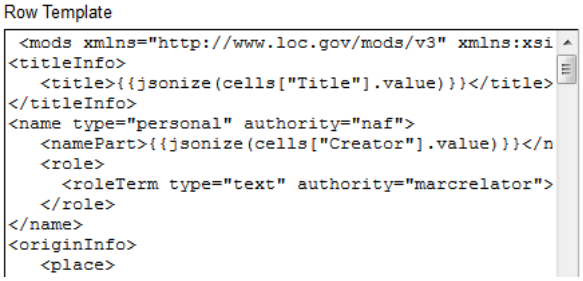
The first two phases of this project involved downloading Web of Science citation data to identify open access works from UNC Charlotte authors that are not in Niner Commons, and then reaching out to said authors to secure approval for us to deposit these works on their behalf. Once faculty and staff have given us their approval, we can batch ingest the works. This involves three major steps: (1) transforming the Web of Science citation data into MODS records that meet the metadata standards of Niner Commons, (2) converting this metadata into XML records, and (3) batch ingesting the resulting XML records and their corresponding PDF files to include the works in Niner Commons.

Process

1. Upload the spreadsheet to OpenRefine as a txt file. Select to **not** store the blank rows.
2. Transform the metadata in OpenRefine using the metadata crosswalk.
3. Rename columns in OpenRefine to perfectly match the metadata spreadsheet
4. Generate an XML file with all MODS records in OpenRefine
    a. Select Export > Templating.
    b. Substitute the template on the left with our own Niner Commons template, as follows:
        i. Prefix: `<?xml version="1.0" encoding="UTF-8"?> <modsCollection xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-7.xsd" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">`



        ii. Row template: the largest middle chunk of the template, except for the very last line

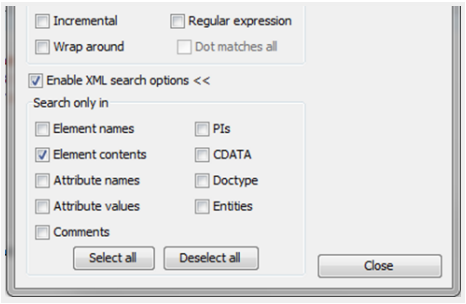


iii. Row Separator: Nothing in row separator

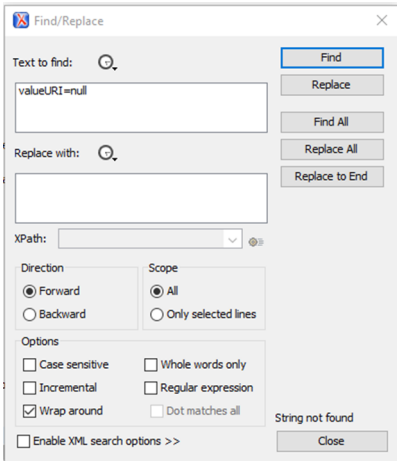
iv. Suffix: `</modsCollection>` (the last line of the template)

c. This process will generate a single XML file containing multiple MODS metadata records. Review the preview box on the right to confirm that data is appearing in the correct elements.
d. Download the text file with multiple MODS records enclosed. Save it with an .xml extension

5. Clean the resulting records in Oxygen XML Editor.
   a. Open the XML file
   b. OpenRefine inserts quotation marks around all content elements. These need to be removed.
      i. Click Find > Find and Replace
      ii. In the Find/Replace box, select Enable XML search options and then Element contents

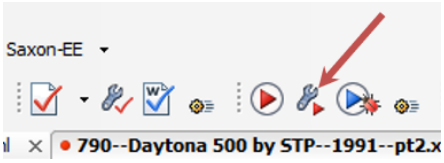


      iii. In “Text to Find”, type a quotation mark. In “Replace with,” place your cursor and leave blank. NOTE: If the cursor is not blinking in this box, the Replace command that you are about to execute won’t function properly.
   c. Next, remove the “valueURI=null” strings from the file following a similar process. Go to the find/replace window, uncheck the Enable XML search options box and put valueURI=null in the “Text to find” area and nothing in the “Replace with” area. Your window should look similar to this:

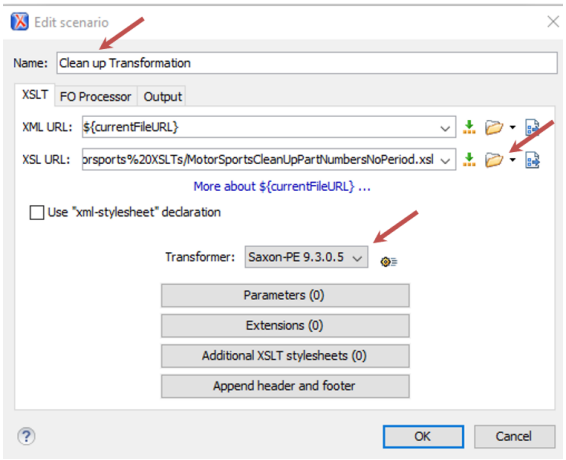


d. And finally, do the same to remove “null” data elements. Go to the find/replace window, keep the Enable XML search options box unchecked, and put >null< in the “Text to find” area and >< in the “Replace with” area.
e. Save the file.

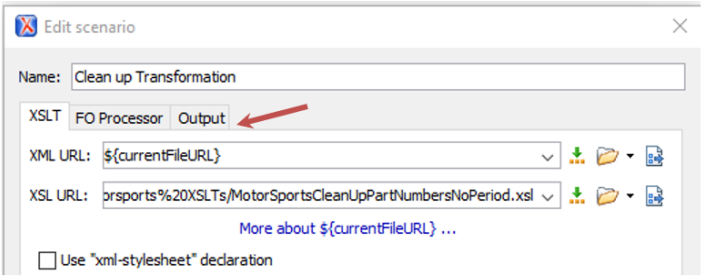
6. Run the first two XSLTs to clean nulls and spacing.
   a. Click on the small wrench icon.



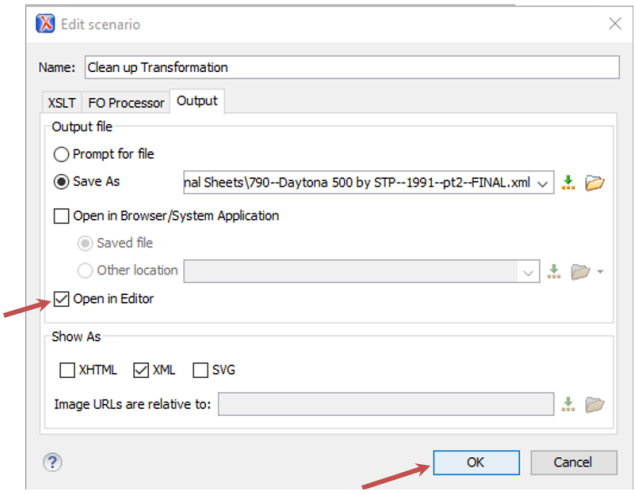
   b. In the “Configure Transformation Scenario” box that will next appear, make sure you select XML transformation with XSLT in the “Scenario type” dropdown list. Then click New. A “New Scenario” box will now appear.
   c. Click the folder icon to the right of the “XSL URL:”. Next, navigate to where you saved the XSLTs locally on your computer. Find the first XSLT (Cleanup1--StripNullNodes.xsl), and then click Open. Now, make sure **Saxon-PE 9.3.0.5**' (The numbers may be different so just make sure Saxon-PE is included) is chosen from the “Transformer” dropdown list. Here you can also name your transformation for future use.



   d. You will now identify the place where you want Oxygen to save the cleaned up file locally on your drive. Click on the Output tab of the “New Scenario” box.



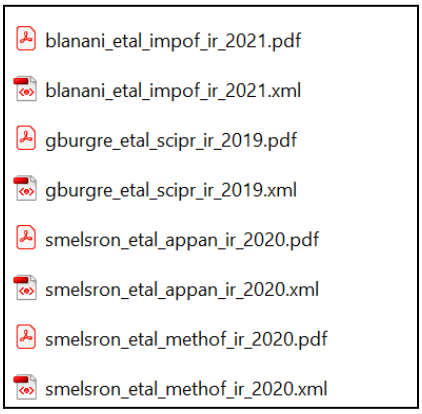
e. Click on the folder icon to the right of the “Save As” box. Navigate to the folder where you will save the cleaned-up file after the XSLT transformation is run. Now, name the file the same as before with “--FINAL.xml” at the end.
f. Before clicking **OK** make sure you check the Open in Editor box, so you can verify the transformation worked and you can initiate the next transformation. Now click **OK**.



g. Click **Transform now**. The XSLT will now run. A green box should appear in the top right if no errors occurred.

h. Run the same process with the second XSLT (Cleanup2.xsl).

7. Run the third XSLT to break up the single XML file containing all of the metadata records into multiple XML files, each having a single metadata record. Follow the same process as you did for the first and second XSLTs.
   a. Once run, you will have a batch of XML records that should match your PDF files.



8. Batch ingest the works into Niner Commons.

a. Copy and paste the XML and PDF files directly into the work-work-ir_works folder on the Timothee shared drive.
b. Within your File Manager, ensure that you can see hidden files within the work-work-ir_works folder. Eliminate any hidden ".dotfiles".
c. Locally on your computer, create a .txt file named “zready.” It must have text in it.
d. Copy and paste your zready .txt file into the work-work-ir_works folder. As soon as you save the zready file into the folder, items will be batch ingested into Islandora. **Never open a zready file that is on Timothee** (it will trigger a reingest).

Many thanks to:

- Joseph Nicholson, Metadata Librarian, for developing and drafting the OpenRefine templating and the OxygenXML Editor steps
- Brad Spry, Software Developer, for developing the batch ingest process