

K@D Metadata Matching on OCLC and ISBN with Python

August 15, 2019

Start with separate, common-delimited text files of the UTL data and your library data
All spreadsheet created will go into the same base folder
[YOUR LIBRARY] is replaced with your library's acronym. Eg. WL for Western Libraries

Set up Python and download scripts (You only have to do this once)

Download oclcxcelmatch.py and isbnexcelmatch.py

If not already installed -

- Install pandas [pip install pandas]
- Install xlrd [pip install xlrd]

Spreadsheet 1. [YOUR LIBRARY] and UTL data

Filename: [YOUR LIBRARY] and UTL data.xlsx

This file will contain:

Sheet 1 - [YOUR LIBRARY] import - raw
Sheet 2 - [YOUR LIBRARY] import - clean
Sheet 3 - UTL import - raw
Sheet 4 - UTL import - clean

Things to do:

- Remove all UTLLIB locations that are E_RESOURCE and other locations as per UTL instructions
- Remove \$b and \$c from the title columns
- Remove extra call numbers (Only keep the first listed. Can use replace ;*)
- **All OCLC and ISBN numbers must have their own row**
*** Many ways to do this using a combination of tools: pivot tables, text to column, insert number of rows, fill row down, shifting cell blocks, merging columns...
- Clean up the OCLC and the ISBN columns, removing all but the number
- Remove leading, trailing, and excessive spaces from OCLC and ISBN rows
- Do not change OCLC or ISBN to text when cleaning or add an apostrophe in front. Instead, you can add a letter to the front of the ISBN

Spreadsheet 2. [YOUR LIBRARY] OCLC Python match

Filename: **[Your Library]1.xlsx**

Things to do:

- Copy '[YOUR LIBRARY] import - clean' to new Excel file
- Headings: BIBNUM, CALLNUM-[YOUR LIBRARY], BARCODE, ITEMNUM, OCLC, TITLE-[YOUR LIBRARY], ISBN-[YOUR LIBRARY]
- Make sure the headings have no leading, trailing, or excessive spaces

Spreadsheet 3. UTL OCLC Python match

Filename: **utl1.xlsx**

Things to do:

- Copy 'UTL import - clean' to new Excel file
- Headings: UTLIB, CALLNUM-UTL, CATKEY, CALLNUM2-UTL, TITLE-UTL, OCLC, ISBN-UTL
- Make sure the headings have no leading, trailing, or excessive spaces
- Remove all blank OCLC rows

Spreadsheet 4. Python OCLC matches

Filename: oclcmatches.xlsx

This file will contain:

Sheet 1 - OCLC MATCHES
Sheet 2 - UNMATCHED

Things to do:

- Open IDLE, file > open > oclcexcelmatch.py
- Change file path to locate your file. Save
- On this file: run> run module
- Open oclcmatches.xlsx
- Create a new sheet called UNMATCHED
- Move all unmatched [Your Library]1.xlsx data to UNMATCHED sheet (These rows will have no CATKEY)

Spreadsheet 5. [YOUR LIBRARY] ISBN Python match

Filename: **[Your Library]2.xlsx**

Things to do:

- Copy oclcmatches.xlsx, UNMATCHED sheet to new excel file
- Headings: BIBNUM, CALLNUM-[YOUR LIBRARY], BARCODE, ITEMNUM, OCLC-[YOUR LIBRARY], TITLE-[YOUR LIBRARY], ISBN
- Make sure the headings have no leading, trailing, or excessive spaces

Spreadsheet 6. UTL ISBN Python match

Filename: **utl2.xlsx**

Things to do:

- Copy UTL import - clean to new Excel file
- Headings: UTLLIB, CALLNUM-UTL, CATKEY, CALLNUM2-UTL, TITLE-UTL, OCLC-UTL, ISBN
- Make sure the headings have no leading, trailing, or excessive spaces
- Remove all blank ISBN rows

Spreadsheet 7. Python isbn matches

Filename: isbnmatches.xlsx

This file will contain:

Sheet 1 - ISBN MATCHES
Sheet 2 - UNMATCHED

Things to do:

- Open IDLE, file > open > isbnexcelmatch.py
- Change file path to locate your file. Save
- On isbnexcelmatch.py: run> run module
- Move unmatched rows to UNMATCHED sheet

Spreadsheet 8. [YOUR LIBRARY]-UTL matches

Filename: [YOUR LIBRARY]-UTL Matches.xlsx

This file will contain:

Sheet 1 - copy oclcmatches.xlsx, save as OCLC
 Sheet 2 - copy isbnmatches.xlsx, save as ISBN
 Sheet 3 - combine OCLC ISBN, save as COMBINED ***These are the matches
 Sheet 4 - from isbnmatches.xlsx, UNMATCHED sheet

Things to do:

- Remove duplicates from Sheet 3 - COMBINED
- Favour Downview records over all other locations when removing duplicates
- Doublecheck that TITLE-[YOUR LIBRARY] and TITLE-UTL match, if they don't, move those rows to UNMATCHED sheet
- Move all unmatched [YOUR LIBRARY] rows from COMBINED sheet to UNMATCHED sheet
- **All ITEMNUM on COMBINED and UNMATCHED sheets must have their own row**
- Ensure numbers match with original [YOUR LIBRARY] data
- Remove any ITEMNUMS not in original [YOUR LIBRARY] data
- Add any ITEMNUMS from original [YOUR LIBRARY] data that are missing to UNMATCHED sheet

Spreadsheet 9. [YOUR LIBRARY]-UTL matches sorted into streams

Filename: [original file name](K@D by stream).xlsx

This file will contain:

Sheet 1 - DISCARD STREAM
 Sheet 2 - STREAM 1 - UTL RECORD
 Sheet 3 - STREAM 2 - NO UTL

Things to do:

- Copy [YOUR LIBRARY]-UTL Matches, UNMATCHED to STREAM 2 - NO UTL
- Copy [YOUR LIBRARY]-UTL Matches, UTLLIB 'contains Downview' to DISCARD STREAM
- Copy [YOUR LIBRARY]-UTL Matches, UTLLIB 'does not contain Downview' to STREAM 1 - UTL RECORD