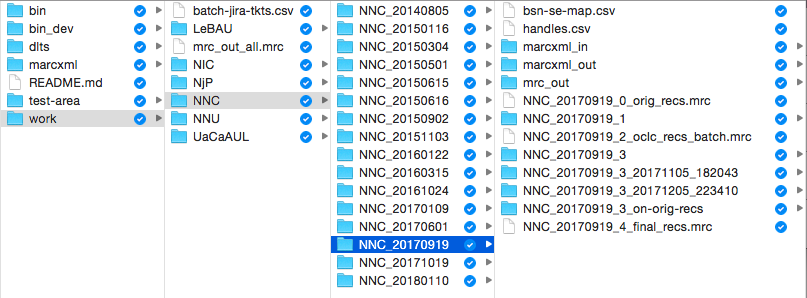
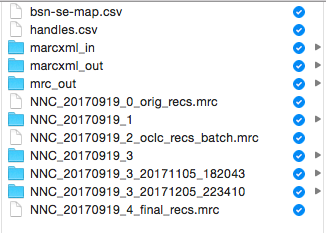
**Procedures for processing MARC files for ACO**

* (Once Only) Set up clone of git repository for "aco-karms" on local computer using git command:

git clone https://github.com/NYULibraries/aco-karms.git

* NOTE: All python scripts are very dependent on correct filenames and batch names. Here is the basic file structure for a completed batch:





Input Received from DLTS

Output from **3rd step**

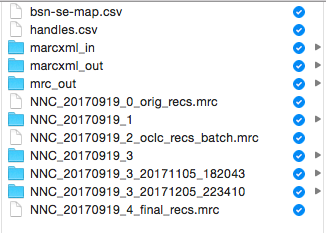
aco-3-process.py

Output from **1st step**

aco-1-xml2mrc-oclc-nums.py

Output from **2nd step**

MarcEdit – OCLC Record Downloader



BSNs 🡨🡪 URL Handles Mapping

BSNs 🡨🡪 Source Entities (Volumes) Mapping

FINAL Sets of INDIVIDUAL Marc Records for DLTS Publishing

OCLC Records

Partner’s ORIGINAL Records

OCLC Nums & Analysis Report

FINAL Set of ALL Marc Records in SINGLE File

1st QC

2nd QC

Last QC

**Processing NEW batches:**

1. Update the master branch on your local git repo clone using git commands:

git fetch THEN git pull origin master

1. Create a new branch on your local git repo clone using git command:

git checkout -b [BATCH NAME] -->

git checkout -b NNC\_20170919

1. Navigate to the "bin" folder: cd bin
2. View the files in bin: ls

----------------------------------------------------------------------------------------------------------------------

1. STEP 1 - Run the Python script to process the MARCXML files and OCLC numbers using command:

python aco-1-xml2mrc-oclc-nums.py

Enter the institutional code:

* Columbia = NNC
* Cornell = NIC
* Princeton = NjP
* New York Univ = NNU
* AUB = LeBAU
* AUC = UaCaAUL

Enter the batch date in the form: YYYYMMDD

NOTE: This script generates the file of original MARC records (named “[BATCH NAME]\_0\_orig\_recs.mrc”) and the folder with the suffix "\_1"

----------------------------------------------------------------------------------------------------------------------

1. STEP 2 - *For Columbia, Cornell, Princeton or AUC records* ***ONLY*** - Process the OCLC numbers using MarcEdit's OCLC Record Downloader to download the corresponding OCLC records
   * NOTE: You must enter the OCLC API settings for NYU's account into MarcEdit
   * Open MarcEdit and navigate to the OCLC Record Downloader
   * For the source file of OCLC numbers, choose the file generated from the previous step found in the folder with the suffix “\_1”, named:

[BATCH NAME]\_1\_oclc\_nums\_for\_export.txt" --> NNC\_20170109\_1\_oclc\_nums\_for\_export.txt

* + For the output file of OCLC records, navigate to the parent folder for the batch and Enter the name of the save file using this pattern:

[BATCH NAME]\_2\_oclc\_recs\_batch.mrc --> NNC\_20170109\_2\_oclc\_recs\_batch.mrc

----------------------------------------------------------------------------------------------------------------------

1. STEP 3 - Run the Python script to check and convert records based on the OCLC records, using command:

python aco-3-process.py

Enter the institutional code:

Columbia = NNC New York Univ = NNU

Cornell = NIC AUB = LeBAU

Princeton = NjP AUC = UaCaAUL

Enter the batch date in the form: YYYYMMDD

NOTE: This script generates the folder with the suffix "\_3”

1. Check the file named "[BATCH NAME]\_3\_all\_recs\_analysis.txt" to see how many records were logged with errors
2. Review the errors report file found in the folder with the suffix “\_3” – For Example: NNC\_20170109 / NNC\_20170109\_3 / NNC\_20170109\_3\_errors\_all.txt

Examples of errors to IGNORE:

* + Key fields missing 880s where the field is NOT Arabic language (e.g., French, English, etc.)
  + Miscellaneous errors for presence of 006 or 007 fields

**UPDATE GIT:**

1. Navigate to the top folder of the local clone of the git repo: cd ..
2. Update github with the feature branch and changes on your local clone, using the following commands:
   * **Check** which files have been changed on branch

git status

* + **Add** the files to be committed

git add work/[INST CODE]/[BRANCH NAME]\* -->

git add work/NNC/NNC\_20170109\*

* + **Commit** the changed/added files

git commit –am ‘Process [NEW/UPDATES] – [INST.BATCH] – [BATCH NAME] – [JIRA TKT #]’

For Example:

git commit -am 'Process new - Columbia-12 - NNC\_20170109 - DLTSACO-464'

\* NOTE: If there are NO errors to be reviewed and fixed, add “ – FINAL SET” to the end of the Commit message

* + **Push** the committed files to github (e.g., git push origin NNC\_20170109)

git push origin [FEATURE BRANCH NAME] -->

1. Go to the aco-karms repo on github.com via a browser and process the pull request, then delete the feature branch on github.com
2. On the command line, Move to the master branch on your local clone using command:

git checkout master

1. Update the master branch on your local clone using commands:

git fetch THEN git pull origin master

1. View branches on local clone using the command:

git branch

1. Delete the feature branch on the local clone using command:

git branch -d [BRANCH NAME]

For Example:

git branch -d NNC\_20170109

**If there ARE errors to be reviewed and fixed:**

1. Copy the 2 files named "[BATCH NAME]\_3\_errors\_all.mrc" and "[BATCH NAME]\_3\_errors\_all.txt" into the shared folder on Box so Adham can retrieve them for QC
2. Assign the Jira ticket to Adham with a message similar to this:

*Adham, This new batch [OR your updates file] has been processed - Columbia-12 (NNC\_20170109) - and is ready for your 1st round of QC. There are 37 records in the errors file, which have been copied to Box. Please let me know if you've any questions or problems.*

*Thanks!*

*Batch: Columbia-12*

*Batch Name: NNC\_20170109*

*QC Round: 1st round*

*No. Records in Errors File: 37*

*For everyone, here are the stats for this 1st round of QC:*

*----------------------------------------------------------*

*Total records processed - batch NNC\_20170109: 48 records*

*Report produced: 2017-01-17 14:04:13.376957*

*Input File (oclc\_batch): 48*

*Input File (orig\_no\_oclc): 0*

*Records where OCLC nums did not match: 0*

*Total records containing any type of error for this round: 37 (77.1%)*

*Total records passing to final version for this round: 11 (22.9%)*

*-----------------------*

*Records containing 880 script fields: 27 (56.3%)*

*Records with NO 880 script fields: 21 (43.8%)*

*Records missing key 880 script fields: 15 (31.3%)*

*Records containing unlinked 880 script fields: 10 (20.8%)*

*Records containing series errors in 490/800/810/811/830 fields: 0 (0.0%)*

*Records containing miscellaneous errors: 8 (16.7%)*

*Records containing bad encoding replacement character: 0*

*Records containing RDA fields: 48 (100.0%)*

*Records with NO 050 or 090 call number fields: 8 (16.7%)*

NOTE: The stats are found at the top of the text file named “BATCH NAME]\_3\_all\_recs\_analysis.txt” in the folder suffixed with “\_3”

1. Once Adham has finished his QC, he will place his updates file onto the shared Box folder at: BoxSync/ACO-QC-Files

**If there are NO errors to be reviewed and fixed:**

1. Assign the Jira ticket to Joe Pawletko (DLTS) and let him know that QC is finished and the batch is ready for the next steps of publishing
2. Go to the aco-karms repo on github.com via a browser, and update the batch queue table in the file found at: aco-karms/work/batch-jira-tkts.csv

If there **ARE** **errors** to be reviewed and fixed:

* + Change the "Cataloging Status" column noting the QC level and number of records in the errors file – e.g., "QC1-37r"
  + Change the "Owner" column to "AA" for Adham Alok

If there are **NO** **errors**:

* + Change the “Cataloging Status” column to “Done”
  + Change the “Owner” column to “n/a”

**Processing QC UPDATES:**

* Retrieve the updates file from the shared ACO folder on the Box drive: BoxSync/ACO-QC-Files

NOTE: It should be named in the pattern: [BATCH NAME]\_3\_updates.mrc -->

NNC\_20170109\_3\_updates.mrc

* Copy the updates file into the folder in the batch that has the suffix "\_3"
* Repeat Steps 1-4 above, Skip Steps 5-6, and Repeat Steps 7-21