Section 1:

Setting up Python environment:

-Download Anaconda | Anaconda Distribution

-Use a python source code editor preferably visual studio code

-Go to start and select Anaconda powershell and type “conda activate base” and then type “code” (this opens up VSC from the command line). This activates conda base and starts VS code using conda. Alternately, you could activate conda following the directions: How to activate conda environment in VS code | by Udi Yosovzon | Medium

-Install git. Go to Git - Downloads (git-scm.com) and download, setup Git with default options.

Section 2:

Cloning the University of Arizona Libraries data curation codes from github:

-In the anaconda powershell terminal type the following commands:

(Some of the commands below are copy pasted from UAL-RE/LD-Cool-P: Python tool to enable data curation (github.com) with some changes)

(Following command creates a folder called “Curation” at the following path: C:\Users\username\Anaconda3\envs)

```
$ conda create -n curation python=3.9.0
```

Click y when Proceed ([y]/n) option pops up

(Following command activates the curation folder)

```
$ conda activate curation
```

(The above command creates a folder curation whose path is at C:\Users\username\Anaconda3\envs\curation)

-Go to File->OpenFolder on VS Code and select curation folder at the above path

-Go to extensions in the left hand corner on VS Code and enter python, select Python IntelliSense (Pylance) and install this python interpreter. After installation, click shift+ctrl+P, type Python interpreter and select Python 3.9.0 (‘curation: conda)

(**Tip**: For quick access to the curation folder, go to File->Preferences->Settings in VSC, type terminal.integrated and change the “python.defaultInterpreterPath”, "terminal.integrated.cwd" and "PYTHONPATH" to "C:\\Users\\username\\Anaconda\\envs\\curation" , replace username with your username, this path might be different on mac OS)

Now, install git (if you haven't already). Go to Git - Downloads (git-scm.com) and click Download, open the .exe file from downloads and setup Git with all the default options.

On VS Code, go to "Terminal" tab and open new terminal, under this terminal on the right column click the drop-down menu, next to "+" tab and select "git bash"

Next, clone the UA Libraries' forked copy of figshare, in the git bash terminal type: git clone https://github.com/UAL-RE/figshare.git

(Alternately, this can be cloned by shift+ctrl+p and pick Git: Clone, go to the github URL above , click code, copy https link and paste it in VSC Git: Clone, press enter, pick curation directory)

Now, run the setup.py script as follows:

(The path below will be C:/Users/username/Anaconda3/envs/curation, where username is what you have to enter, alternately you could go to the folder where curation folder was created, and copy paste the path below)

```
(curation) $ cd /path/to/parent/folder
```

(The path below will be C:/Users/username/Anaconda3/envs/curation/figshare, where username is what you have to enter, alternately you could go to the folder where curation folder was created and copy paste the path below)

```
(curation) $ cd /path/to/parent/folder/figshare
```

```
(curation) $ (sudo) python setup.py develop
```

(**Errors:** If setup doesn't develop or conda is not recognized then activate it from the anaconda powershell, (Start->anaconda powershell), type conda activate base, and open vs code with command "code" from the anaconda powershell, python - Error when trying to use conda on vs code: conda : The term 'conda' is not recognized as the name of a cmdlet - Stack Overflow)

Next, get out of the figshare folder and back to the curation folder (cd ..). Clone this repository (LD-Dool-P) to the curation folder and install with the setup.py script:

(The path below will be C:/Users/username/Anaconda3/envs/curation/, where "username" is to be replaced based on the path it appears on your folder, alternately you could go to the curation folder and copy paste the path below)

```
(curation) $ cd /path/to/parent/folder
```

```
(curation) $ git clone https://github.com/UAL-RE/LD-Cool-P.git
```

(The path below will be C:/Users/username/Anaconda3/envs/curation/LD-Cool-P (cd LD-Cool-P), where username is to be replaced, alternately you could go to the folder where LD-Cool-P folder was created, and copy paste the path below or type cd LD-Cool-P to go inside the curation directory)

```
(curation) $ cd /path/to/parent/folder/LD-Cool-P
```

```
(curation) $ (sudo) python setup.py develop
```

This will automatically install the required pandas, requests, numpy, jinja2, tabulate, and html2text packages.

You can confirm installation via conda list

```
(curation) $ conda list ldcoolp
```

You should see that the version is 1.1.8

Section 3:

Cloning the Virginia Tech Libraries data curation codes from github:

-Clone VT Figshare repository :

Open bash on VS Code, clone VT repository in the curation folder using the following command:

```
$ cd C:/Users/username/Anaconda3/envs/curation
```

```
$ git clone https://github.com/PadmaCarstens/VTechDataRepo.git
```

Section 4:

Setting up configurations on your local computer:

Go to File->Open File-> Open generate_config_example.py located at Users/username/Anaconda3/envs/curation/VTechDataRepo/generate_config_example.py, go to File->Save As-> (Go to curation/VTechDataRepo folder) generate_config.py

Open .gitignore and make sure generate_config.py appears in this list.

Go to generate_config.py and fill in all the folder paths and credentials. The third tag needs to be filled in each setting. For example:

"FigshareSettings", "FigshareArticleID", "XYZ" then "XYZ" is the setting to be filled in by the user. Fill in the Figshare Article ID XYZ as the number that appears under "Cite" red button for the article in review/published.

Replace the root name and/or path to what appears on your local computer. For example: replace "padma" and the path with what appears on your local computer path.

Note: The config path "LargeBagsPath"," F:/VTechbags" is for downloading huge datasets to a different location (in this case to Sandisk), this option can be ignored since it requires uncommenting a section of the script in the IngFolder_Download_TransferBagAPTrust.py and PubFolder_Download.py in order to enable downloading to a different path.

After these configurations are saved, run generate_config.py, this creates configurations.ini file (this is hidden if scripts are pushed to github through .gitignore)

Section 5:

Replacing UAL script and figshare script with the customized VT script for curation:

-Copy the script retrieve.py from the following folder:

C:\Users\username\Anaconda3\envs\curation\VTechDataRepo\Figshare-LDCoolP

and paste ("Replace the file in destination"/overwrite the original retrieve.py") it in the following folder:

C:\Users\username\Anaconda3\envs\curation\LD-Cool-P\ldcoolp\curation

-Copy the script figshare.py from the following folder:

C:\Users\username\Anaconda3\envs\curation\VTechDataRepo\Figshare-LDCoolP

and paste ("Replace the file in destination"/overwrite the original figshare.py") it in the following folder:

C:\Users\username\Anaconda3\envs\curation\figshare\figshare

Copy client_secret.json from Curation Workflow folder and paste it at C:\Users\padma\Anaconda3\envs\curation

Section 6:

Filling ingest/published article information in VTDR v7 spreadsheet and setting up DART:

Before running these scripts, the following steps need to be followed:

1. Please make sure the ingest/published record information for the article in review/published is entered in the VTDR V7 spreadsheet in the corresponding sheet “Ingest” or “Published”
2. Please make sure DART app is installed and set up for uploading bags to APTrust following instructions in CurationWorkflow_SetupDART_APTrustDeposit.docx in Curation Workflow folder

Section 7:
Running the VTDR Workflow scripts:

-Creating README.rtf:

-Enter the article id whose README.rtf needs to be created in generate_config.py and run generate_config.py
-Make sure ingest information in the “Ingest” sheet in the VTDR V7 spreadsheet is filled in as described in section 6
-Run the script AutomatedREADMErtf.py located at Curation/VTechDataRepo/Figshare-APTrust
-A README.rtf will be created at Curation/README_FILES_timestamp_authorname

-Creating Ingest bag creation and depositing it to APTrust/ VT S3:

1 Open the script: IngFolder_Download_TransferBagAPTrust.py
2 Pick a workflow for depositing bag to APTrust demo/repo/VT S3 following the instructions in section 8.
3 Run the script IngFolder_Download_TransferBagAPTrust.py
Running this script after picking the workflow above will create an ingest folder in the “Curation” folder which looks like VTDR_I00NBR_lastnamefirstnameinitial_lastnamefirstnameinitial_v0X_date. An ingest bag (ingest folder in tar format and with tag values in it) will also be created at Username/.dart/bags or at the “Output Path” picked in “Application Setting” in DART app. The bag name will have the same naming convention but with a .tar at the end. This bag will be uploaded to aptrust demo/repo/VTs3 bucket depending on the workflow selected in step 2. The bags on demo or repo can be checked for upload at demo.aptrust.org or repo.aptrust.org

-Steps for Publication bag creation and deposition to APTrust:

1 Open and run the script PubFolder_Download.py, this script downloads the published dataset and creates a publication folder in the “Curation” folder

2. Open the publication folder created, this folder will have a naming convention like: VTDR_P00XYZ_I00XYZ_DOI_XYZ_lastnamefirstinitial_v0X_date.
3. Fill in the Provenance Log and save Email interactions and save them as ProvenanceLog.rtf and Email_Correspondence (or Email_Correspondence1,Email_Correspondence2 etc in case of multiple email threads) in VTCurationServicesActions folder found at the path: C:\Users\username\anaconda3\envs\curation\VTDR_P00XYZ_I00XYZ_DOI_XYZ_lastnamefirstinitial_v0X_date\VTCurationServicesActions
4. Open PubBagDART_TransferBagAPTrust.py
5. Pick a workflow for depositing bag to APTrust demo/repo/VT S3 following the instructions in section 8.
6. Run the script PubBagDART_TransferBagAPTrust.py

   This will create a publication bag (publication folder in tar format and with tag values in it) “Username/.dart/bags” and has the same naming convention but with a .tar at the end. This bag will be uploaded to aptrust demo/repo/VTs3 bucket based on what was selected in step 5. The bags on demo or repo can be checked for upload at demo.aptrust.org or repo.aptrust.org

7. Check the bag created following the instructions in section 9

Section 8:

Enabling/disabling a workflow:

For deposition to the **APTrust demo bucket**, uncomment(remove “#”) the following line in the script:

```
job = Job("APTrust Demo Workflow for Virginia Tech",aptrustBagName)
```

and comment (add “#”) the following lines (lines 119 and 121)in the script:

```
#job = Job("APTrust Production Workflow for Virginia Tech",aptrustBagName)
#job = Job("VT Workflow for depositing bag to VT library S3 bucket",aptrustBagName)
```

For deposition to the **APTrust repo bucket**, uncomment(remove “#”) the following line in the script:

```
job = Job("APTrust Production Workflow for Virginia Tech",aptrustBagName)
```

and comment (add “#”) the following lines (lines 117 and 121) in the script:

```
#job = Job("APTrust Demo Workflow for Virginia Tech",aptrustBagName)
#job = Job("VT Workflow for depositing bag to VT library S3 bucket",aptrustBagName)
```

For deposition to the **Virginia Tech library S3 bucket**, uncomment (remove “#”) the following line in the script:

```
job = Job("VT Workflow for depositing bag to VT library S3 bucket",aptrustBagName)
```

and comment (add “#”) the following lines (lines 117 and 119) in the script:

```
#job = Job("APTrust Demo Workflow for Virginia Tech",aptrustBagName)
```

Section 9:

Checking the bags created by DART:

To check the contents of the bag in tar format created at the output path, go to the Start->Command Prompt and type:

```
>cd <Output Path> (For Example: if bag is created at an output path
C:\Users\username\Documents\DART then change directory to

cd C:\Users\username\Documents\DART)

>tar -xvf VTDR_I00NNN_XYZ_XYZ_v01_xof8_date.tar
```

This will extract the contents of the bag above in the same directory as the bag in tar format

Section 10:

Possible errors:

Error:

```
“File "C:\Users\username\Anaconda3\lib\site-packages\redata-0.4.1-
        py3.9.egg\redata\commons\logger.py", line 3, in <module>

from os import path, uname, chmod, mkdir

ImportError: cannot import name 'uname' from 'os' (C:\Users\padma\Anaconda3\lib\os.py)”
```

Possible solution:

Then, open logger.py in the folder above and replace

“from os import path, uname, chmod, mkdir” on line 3 with the following 2 lines:

```
from os import path, chmod, mkdir

from platform import uname
```

Now, save the logger.py and run again

Error:

“Module gspread not found”

Then install gspread:

```
-pip install gspread
```

Error:

“ModuleNotFoundError: No module named 'oauth2client'”:

Then, install oath2client:

```
-pip install oauth2client
```

Error:

“ModuleNotFoundError: No module named 'PyRTF'”

Then, install PyRTF:

```
-pip install PyRTF
```

Error:

ModuleNotFoundError: No module named 'bagit'

Then, install bagit:

```
-pip install bagit
```

Error:

ModuleNotFoundError: No module named 'rdflib'

Then, install rdflib:

```
-pip install rdflib
```