**HBCU Data Processing Documentation (Refer HBCU Dashboard Files)**

* All the Python scripts for processing Data are saved in the Folder named “Programs\_Automation\_Scripts“.
* In the first Process, all the 240 files would need to be downloaded from Scival.
* The below steps were implemented for Generating the consolidated data:

1. Search for HBCUs in the ScviVal search box.

(You will get a list of 80 Universities)

1. Select an individual university.
2. Now Go the Authors section and export the data as csv.
3. Then Go to “Collaboration“ -> “Top Collaborating Institutions“. Export the data as csv
4. Then go to “Published“ -> “by Subject Area“. Export the data as csv.
5. Now all of these files would be on you PC in the downloads section.
6. Make a folder called PrepareFiles and paste all the 3 Files in it.
7. Now from the Programs Automation Scripts, paste the python code named "Renaming\_University\_Files".
8. Open cmd in the folder.
9. Run the program through the command prompt. Type "python (Name of the File)".
10. Now the files would be renamed. Once this is done, you can move these files to 3 folders namely – “Authors”, “Collaborations” and “Published\_Subject\_Areas”.
11. Repeat process 3 to 11 , 80 times. In this case it is 80, since there were total 80 Universities.
12. Now, you have 3 folders with 80 files each.
13. Go to the folder -> “Authors”
14. From the Program Automation Scripts -> Authors Folder - > copy paste the program – “1\_Modify\_Columns\_1\_Authors” into the Authors folder.
15. Run this program using cmd with the same steps as -10
16. Now 80 New modified files will be created. Delete the previous 80 files. Now these 80 new files will have the correct column names.
17. Now copy paste the program – “2\_Merge\_All\_Files\_Authors” into the Authors folder from Program Automation Scripts -> Authors.
18. Run this program using cmd.
19. Now a new file will be created which will have all the data merged into one. This is your final file.
20. Now the process is similar for the other 3 attributes.
21. Go to the folder -> “Collaborations”. This has all the 80 collaborations part csv files.
22. From the Program Automation Scripts -> Collaboration Folder - > copy paste the program – “1\_Modify\_Columns\_1\_Collaborations” into the Collaborations folder.
23. Run this program using cmd with the same steps as -10
24. Now 80 New modified files will be created. Delete the previous 80 files. Now these 80 new files will have the correct column names.
25. Now copy paste the program – “2\_Merge\_All\_Files\_Collaborations” into the Authors folder from Program Automation Scripts -> Collaboration.
26. Run this program using cmd.
27. Now a new file will be created which will have all the data merged into one. This is your final file.
28. Go to the folder -> “Publication\_SubjectAreas”. This has all the 80 files of Publications – Subject Areas
29. From the Program Automation Scripts -> Publication\_SubjectAreas Folder - > copy paste the program – “1\_Modify\_Columns\_1\_ Publication\_SubjectAreas” into the Publication\_SubjectAreas folder.
30. Run this program using cmd with the same steps as -10
31. Now 80 New modified files will be created. Delete the previous 80 files. Now these 80 new files will have the correct column names.
32. Now copy paste the program – “2\_Merge\_All\_Files\_ Publications” into the Publication\_SubjectAreas folder from Program Automation Scripts -> Publication\_SubjectAreas.
33. Run this program using cmd.
34. Now a new file will be created which will have all the data merged into one. This is your final file.
35. Thus you get 3 final files with all the consolidated Data