

4/30/2020

This document outlines the steps to transform the data from files delimited by "|" to files delimited by ",". This is needed as R is not able to read data files with multi-character separators. Commas are also not usable as separators in this circumstance as there are several commas found inside data fields themselves. Due to the size of the data files, simple find & replace methodology was not effective.

Running this R script effectively calls four Python scripts to transform the "|" to "," in academic, membership, individual\_info and events data files respectively.

While this is a feasible workaround, we recommend that UMAA use files delimited by "," moving forward. If this file format is adopted, the "0 - Python files in R.R" can be removed.

### Scripts to run

## Scripts to run

---

As a one-time set-up, do the following 3 steps:

2. Install reticulate package in R
  - o Run `install.packages("reticulate")` in R
3. Install miniconda in R
4. Run `pip install pandas` in R terminal

After the above 3 steps have been completed, we can execute the code below.

We first set the working directory

```
## set working directory
setwd('D:/Group Folder/SEQUENTIAL FILES')
```

We then run the code below.

```
## Libraries used
library(reticulate)

## Academic csv transformation
py_run_file('0 - Code/Read academic.py')

## Membership csv transformation
py_run_file('0 - Code/Read membership.py')

## Individual_info csv transformation
py_run_file('0 - Code/Read individual_info.py')

## Events csv transformation
py_run_file('0 - Code/Read events.py')
```

New csv files for Academic, Events, Individual\_info and Membership have been created in `1 Data Initial Files` folder by the name of `academic`, `events`, `individual_info` and `membership` respectively.