

UK Longitudinal Linkage Collaboration: File Checker User Guide

Introduction

The File checkers are designed as a user tools to ensure file submissions are in line with the requirements set out in “UK Longitudinal Linkage Collaboration: Guidance for Depositing Data into the UK LLC v1.9”. Each checker verifies that field names and values are of expected syntax and data type. It makes no pass judgement on the contents of the files other than their legality. The checkers take .csv files as input and will output a .txt file detailing any issues that were found.

Installation (executable)

Once you have downloaded the “UK_LLC_File_Checker.zip” file. We recommend extracting it its own private folder because the checker will produce output text files and save them in the same directory that it is run from. For this reason, make sure you save it in a location that allows programs to create and modify files. When extracted, the directory should contain the files shown below.

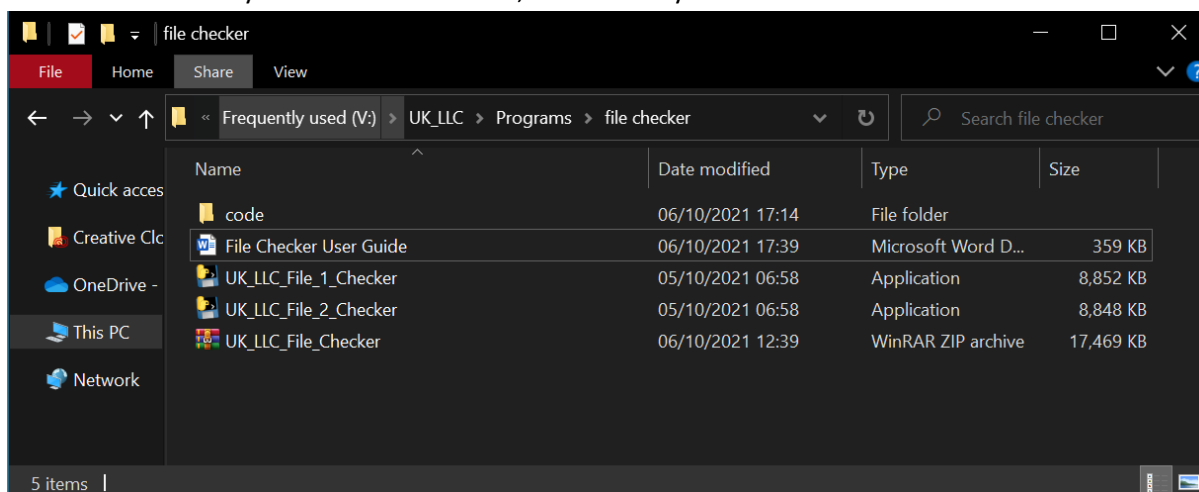


Figure 1. Save Directory

Running

You can execute a checker through the .exe application files by double clicking, or by right clicking and selecting “run”. Anti-viruses tend to be cautious of .exe files, so you may have to explicitly allow it permission to run. If this is not possible you can instead execute the source code provided you have Python 3 installed. To do so, open a command prompt and navigate to the code folder in the save directory. Once there, type ‘python UK_LLC_File_*_Checker.py’ to start a checker, exchanging the Asterix for 1 or 2 as appropriate.

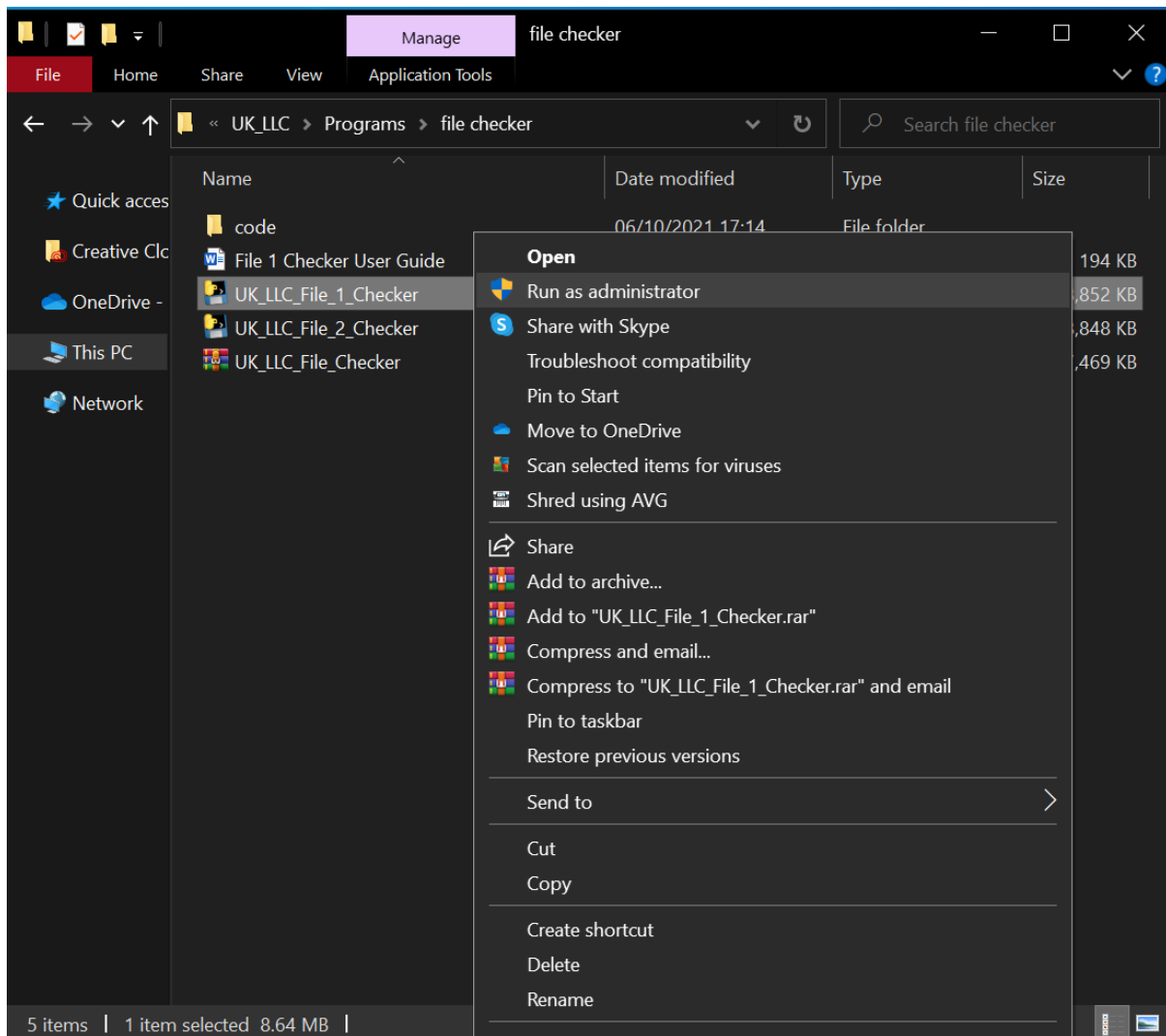


Figure 2. Running from executable



Figure 3. Running from source code

Immediately a console output window and file dialog window should appear. You can use the file dialog to select your file. The dialog only shows .csv files by default, but this can be changed to .txt files with the drop-down box in the bottom right. The program will not allow you to pick files with extensions other than .txt or .csv.

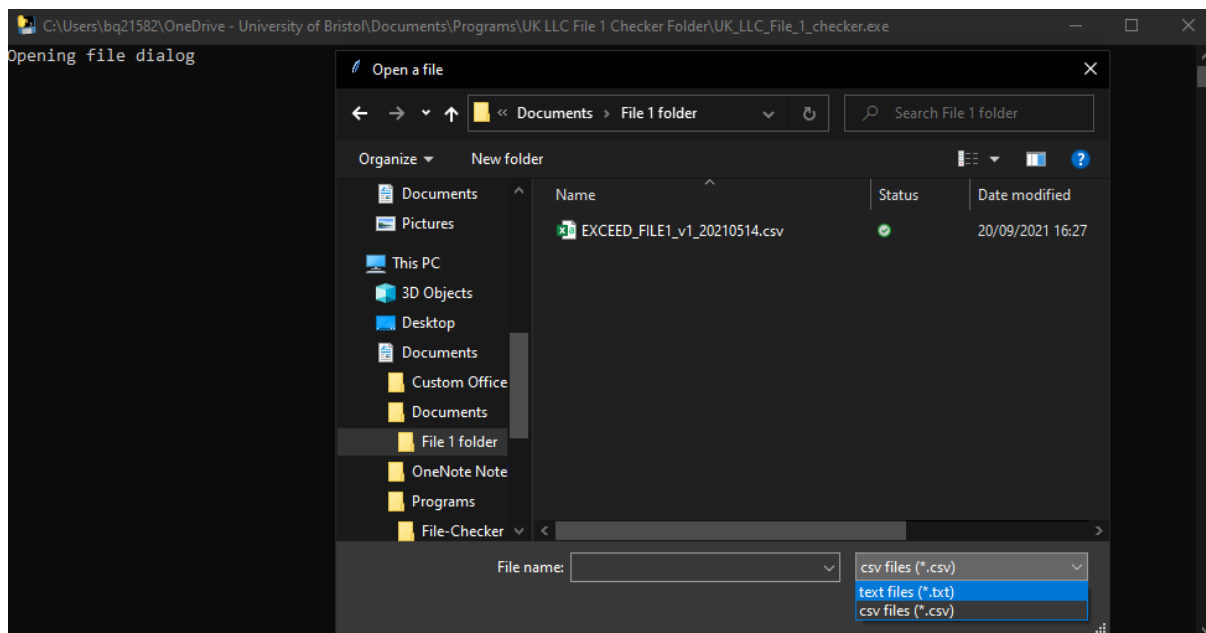


Figure 4. Changing file extension filter

Selecting a file will import it into the program and checks will begin immediately. This process might take several minutes for particularly large files. Progress updates will be printed to the console output window as the checks progress. The console output window will automatically close when the checks are complete.

Output

The checker program will output a log after completion as a .txt in same directory as the checker was run from. The output file will be named in the format:

`'[input_file_name]_Output_Log[time:HHMMSS].txt'`

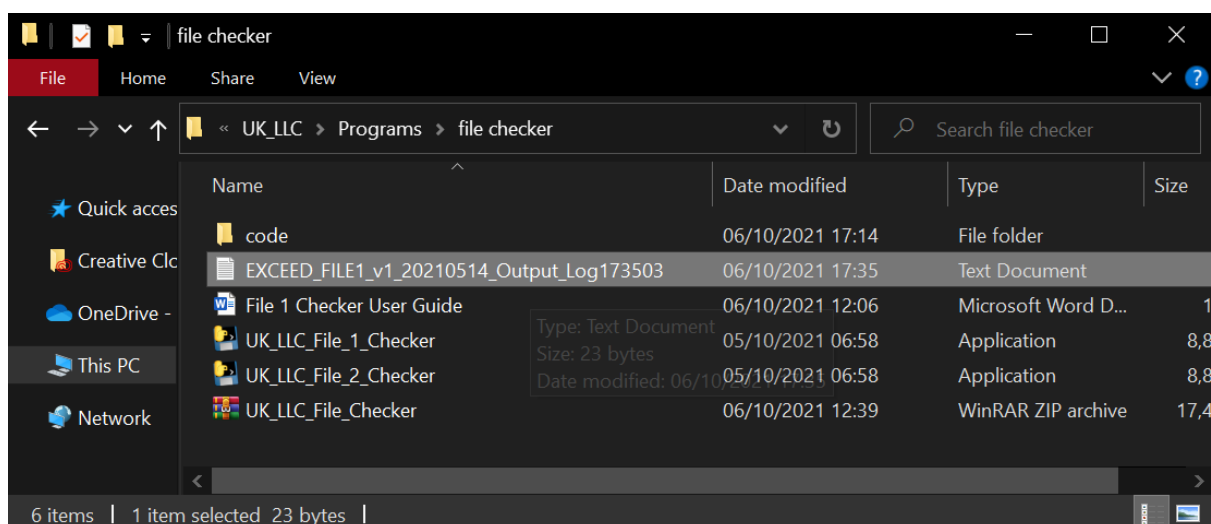


Figure 5. Output file in save directory

The file will contain a list of all errors encountered, including a small description and problematic lines where appropriate. If more than 10 lines suffer from the same error, only the first 10 lines will be listed for the sake of brevity, with the understanding that the error is likely widespread throughout the file.

Bug Fixing and Extensions

These file checkers have been tested on as many foreseeable scenarios as possible, but this is likely not exhaustive. If the checkers falsely identify an error or do not identify an error where you believe to be one, please relay a description of the problem to matt.crane@bristol.ac.uk.

If you have any suggestions for future extensions or ease of use improvements, equally please do not hesitate to contact us.