UK **Longitudinal Linkage Collaboration**

UK Longitudinal Linkage Collaboration
Population Health Sciences
Bristol Medical School
Canynge Hall
39 Whatley Road
Bristol BS8 2PS

The UK Longitudinal Linkage Collaboration (UK LLC)

File 1 Checker User Guide

PUBLIC
Version 2.0
2ⁿᵈ November 2022

Introduction

The File checker is a user tool for verifying the contents of a File 1 submission are in line with the requirements set out in "UK LLC File Formatting Guidance for Depositing Data in the UK LLC Resource v1.15". The checker verifies that field names and values are of expected syntax and data type. It makes no judgement on the contents of the files other than their legality under formatting rules. It is very important that File 1s are cleared through the checker to avoid complications loading them into the UK LLC databank.

The checker includes a section for File 1 Documentation that should be sent to the UK LLC. Information is automatically loaded from a File 1 where possible and cross referenced for correctness. The documentation is saved in a standard JSON format ready to be sent to the UK LLC.

The program is available as a graphical user interface that can be run as an executable or with python. The code is available at https://github.com/UKLLC/File-Checker.

Installation

Go to the UK LLC GitHub at https://github.com/UKLLC/File-Checker and download the code as a zip file. Extract the contents into its own folder. The checker will produce output files in the 'outputs' folder. Please make sure the directory has read and write permissions for this to work. When extracted, the directory should contain the files shown below

[screenshot: file directory]

Running

Run the checker program by double clicking, or right clicking and selecting "run", the [filename].exe application. This will launch a graphical user interface. If you are unable to run executable files, you can instead run the application with Python. You will need an installation of Python 3. Open a command prompt window and navigate to the code folder in the checker's directory. Run the application by typing 'python [filename].py'.

When the interface appears, start by loading a File 1 csv by clicking the 'Load File 1' button. To begin the file checking process, click the 'Start' button. The checks might take several minutes depending on the size of the file. The results of the file checker will be presented in the text field directly below the start button and output as a text file in the outputs folder in the format: *'[input_file_name]_Output_Log[time:HHMMSS].txt'* .

The output is a list of all errors encountered, including a small description of the error and lines of the file where it was encountered 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.

After the automated file checks are completed, fill out the 'File 1 Documentation' section. The date, file name, study name, row count and included participants will be automatically filled from the loaded file where possible. Make sure the number of exclusions and inclusions add up to the total number of participants in the study.