CAP App User Guide

January 2026

Contents

1. Introduction

The CAP app is a tool for users to calculate the disclosure risk of a synthetic dataset. This was developed in the Enhancing Data Accessibility and Security through Data Synthesis (EDASIDA) project.

This document provides information on how to install and run the app. For further detail about the specifics of the CAP calculation see Taub & Elliot (2019)

2. Installation

2.1 Download and extract the files

To begin, download the provided .zip file (available here) and extract its contents to a convenient location on your device. This may take a couple of minutes. After extracting the zip file, open the extracted folder and you should find the application folder ("EDASIDACAPGUI"), together with a "LICENSE" file and "quickstart" files.

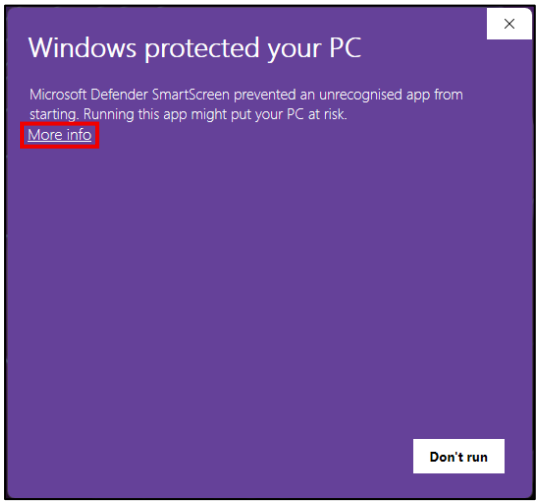
2.2 Launch the application

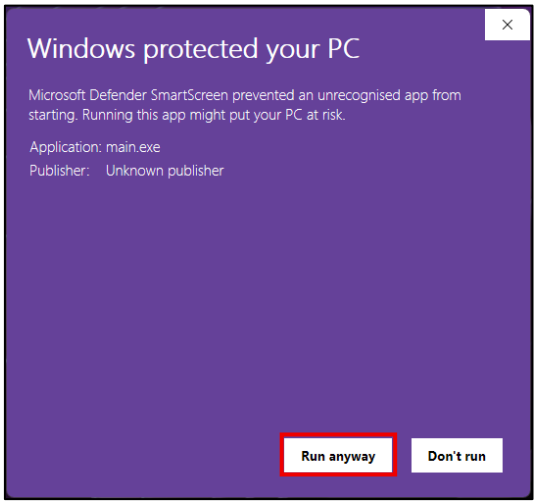
The following instructions are also available in the "quickstart.pdf" file located in the extracted folder.

1. Navigate to the "EDASIDACAPGUI" folder which is inside the extracted folder
2. Locate the file named "main.exe"
3. Double-click "main.exe" to run the application

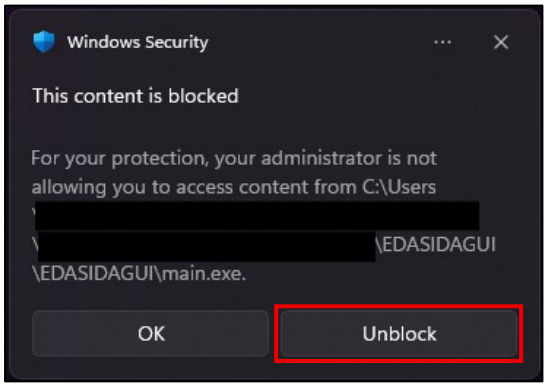
4. If you are prompted with a security warning, please unblock the executable before retrying

2.2.1 Handling Windows security notifications

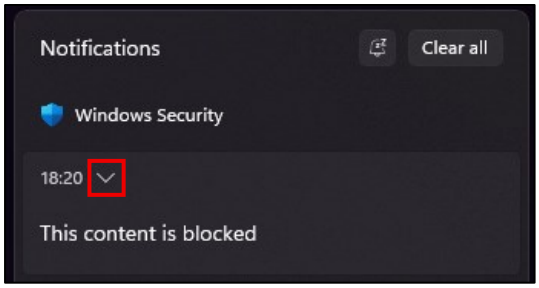


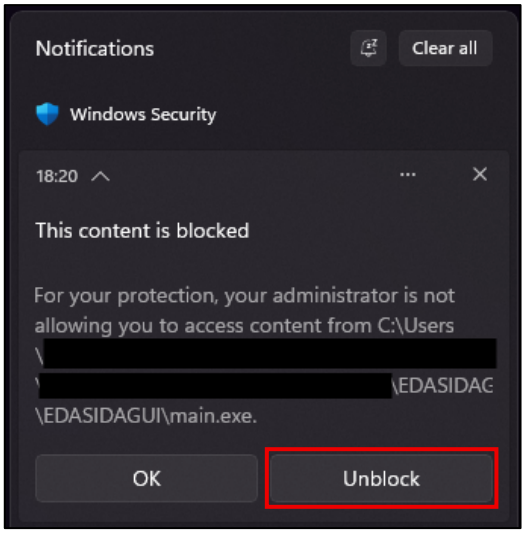


Windows Defender may pop up with a prompt window suggesting it is unsafe to run this application, this can be bypassed by clicking “More info” and then the “Run anyway” button.



If you encounter a Windows Security notification like the one shown above, it means that your device’s security settings are preventing access to the GUI. If you are happy to do so, please click “Unblock” or “Dismiss” when prompted.

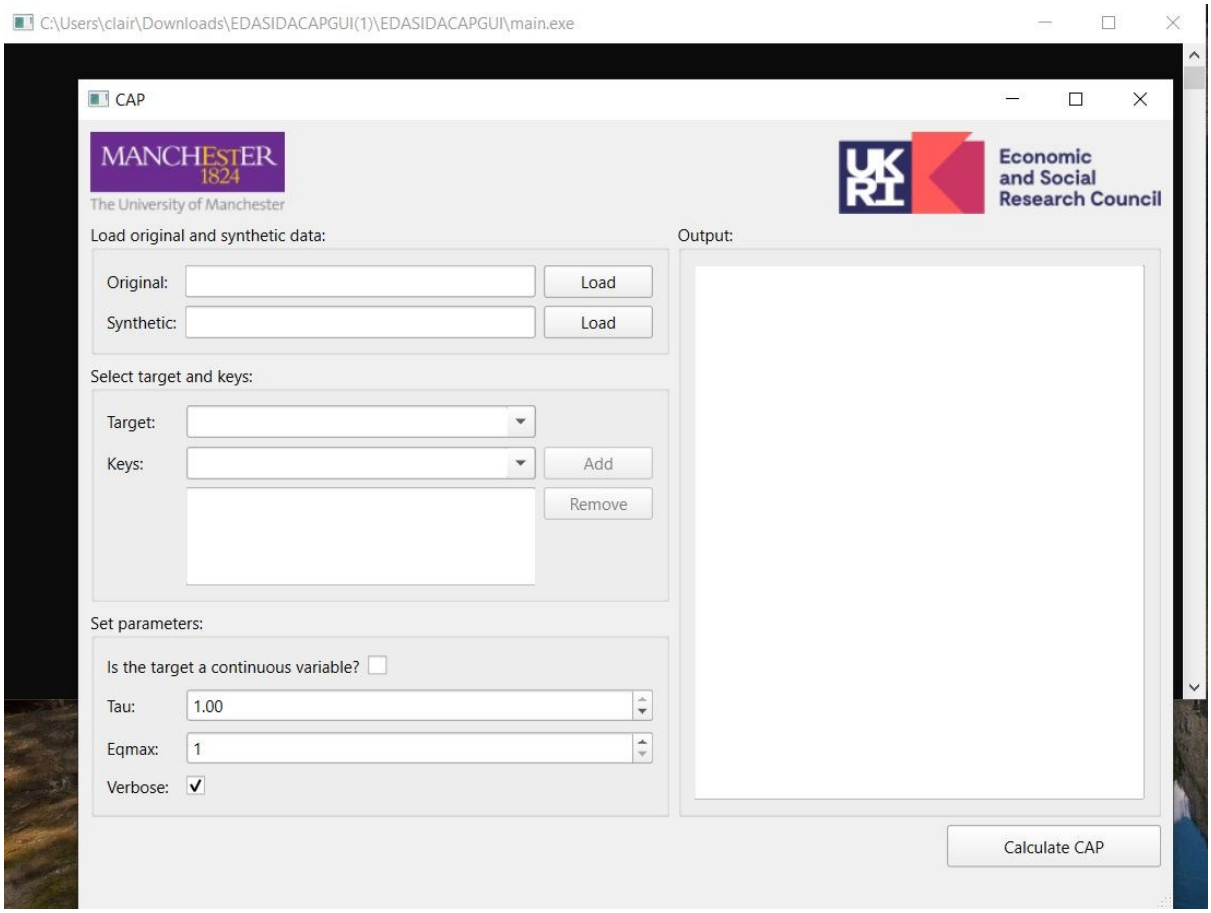




If you could not follow the prompt in the short timeframe of this pop up, you may also find it in the notification centre. Double click on the "main.exe" executable again (if required), to launch the application.

Note: All screenshots in this section were captured on Windows 11.

Once launched, the app should appear alongside a black box window (this console will be used to display dialogue from the app):



3. Using the app

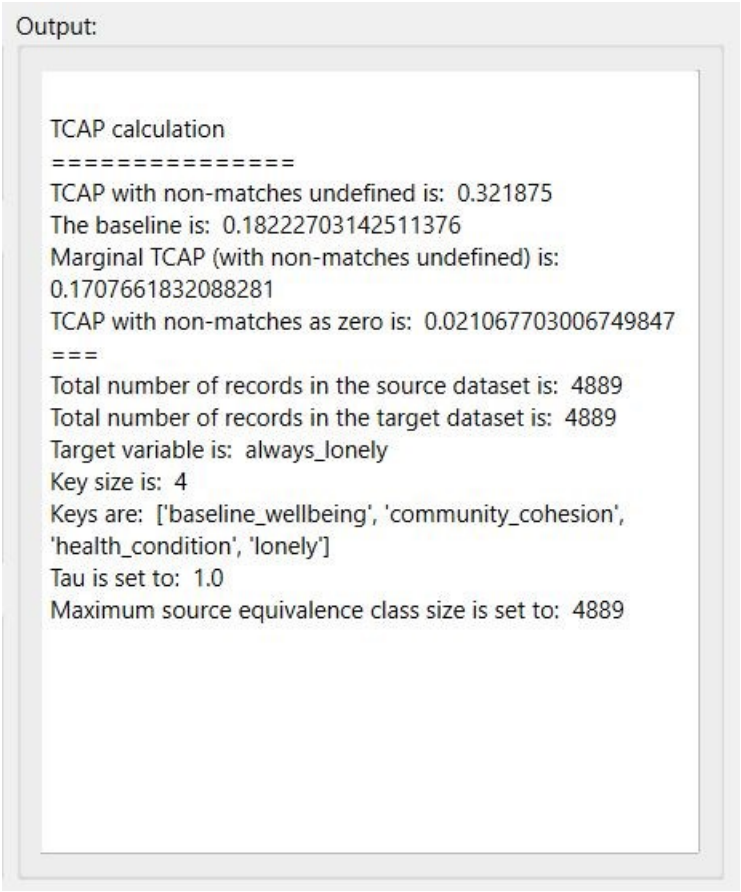
3.1 Calculating the TCAP score

1. Load the original dataset by clicking "Load" next to the "Original" box – a file explorer box will appear and you can locate and load the original .csv file. Only .csv files can be loaded.
2. Load the synthetic dataset by clicking "Load" next to the "Synthetic" box – a file explorer box will appear and you can locate and load the synthetic .csv file. Only .csv files can be loaded.
3. Select a "Target" variable (this should be a categorical variable) by clicking in the drop-down box next to "Target" – a list of all the variable should appear and you can select the relevant variable.

4. Select some key variables (a minimum of three usually makes sense) by clicking in the drop-down next to “Keys”. A list of variables should appear, select one and then click “Add” – it will appear in the box below. Repeat this until all key variables have been added.
5. Click on the “Calculate CAP” button to calculate the score - using the existing parameters (tau=1) will calculate the TCAP score
6. The TCAP score along with various information will appear in the “Output” window on the right
7. To exit the app simply click the X in the top right corner

3.2 Understanding the Output

The Output window will display the disclosure risk score alongside other information about the data. For further information about the TCAP calculation see Taub & Elliot (2019)



- “TCAP with non-matches undefined” is usually the disclosure risk score that most people would require. This is a score between 0 and 1, with 1 indicating most risk.
- The “baseline” score gives the risk score that you would get if just guessing. A TCAP score above that indicates some level of risk.
- The “Marginal TCAP” scales the score between the baseline and 1
- “TCAP with non-matches as zero” treats all non-matches as zero rather than undefined
- The rest of the text summarizes the number of records in the datasets and the variables used as keys and target.

3.3 Experimental output

There is an option to use a continuous variable as the target – this is experimental and a work in progress and should be used with caution.

Simply select a continuous variable as the Target and then tick the "Is the target a continuous variable?" box. Once the calculate CAP button is clicked a single score is provided, along with the summary of the data and variables.

References

Taub J., Elliot M. 2019. "The Synthetic Data Challenge" *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, The Hague, Netherlands*, October 29–31. https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2019/mtg1/SDC2019_S3_UK_Synthethic_Data_Challenge_Elliot_AD.pdf