

Data Protocols

This document briefly describes the main data safety protocols.

We will store different versions of the data, and use them for different purposes. In particular, there will be three versions of the data: The original version with personally identifiable information (PII), the full data set without PII, and small samples (1-10%) of the data set without PII.

Original version with Personally Identifiable Information (PII)

Three copies of the original data with PII information will be kept in portable hard drives, not in ITAM's servers. These hard drives will be encrypted using BitLocker and stored in a locked cabinet (safe). The principal investigators will hold these hard drives.

This data will only be un-encrypted and used if we must use PII to merge these files with other data sources (e.g., using the CURP to merge our data with tax records from SAT). Only principal investigators will have access to the PII data. No other person (research assistant or otherwise) will have access to the PII version of the data.

Full data set without PII

Another copy of the data, without PII, will be kept in a server at ITAM. This data will only be used to estimate the statistical models that will appear in the final version of the papers. This data will never leave this server.

Working samples

For each project, researchers are in charge of authorizing the removal of a sample of the full data set from the server for each project using the data described above. Thus, this data will not have PII. These samples will never have more than 10% of the total number of observations. These samples will be kept in Dropbox and are meant to be used for exploring the data and writing the codes that will be used to estimate the statistical models in the server at ITAM. Only researchers and the people they authorize will have access to this data. However, these data can be shared with researchers outside ITAM.

Data Sharing

Only the working samples of the data will be shared with other researchers. Either researchers have to authorize any data sharing before it happens.

Any researcher that wants to use any version of the data (including research assistants, other ITAM employees, or non-ITAM employees) must sign an agreement with ITAM in which they agree to: a) only use the data for the publication of academic papers with at least one of the ITAM researchers as coauthors, b) only use the data for the project they were given authorization; c) never share the data with other parties without previous authorization. d) never to attempt to identify individuals from the data,

e) report any misuse of the data, including the possibility that subjects may be identified. They must almost sign the standard confidentiality agreement ITAM has.