**RECOMMENDED DATA CITATION:**

This dataset already came with a citation that fits extremely well with the recommended structures. It is listed below:

-MPO Region to County Crosswalk (Halter G, Hitchcock A, Kelly T, Mortenson M. 2010 California Regional Progress Report: One State, Many Regions, Our Future: Tracking Progress Toward Sustainability. Sacramento, CA: Applied Development Economics Inc., Collaborative Economics Inc., Information Center for the Environment at UC-Davis, CALTRANS: Division of Transportation Planning; 2010. [http://dot.ca.gov/hq/tpp/offices/orip/Collaborative%20Planning/Files/CARegionalProgress\_2-1-2011.pdf. Accessed on 7/5/12](http://dot.ca.gov/hq/tpp/offices/orip/Collaborative%20Planning/Files/CARegionalProgress_2-1-2011.pdf.%20Accessed%20on%207/5/12).)

**CONSIDERATIONS FOR LONG TERM PRESERVATION**

During the course of working on this project, from the time that this dataset was first discovered at the end of January 2023 to February 15th, 2023, it has since been removed from the original repository that it was originally found: data.gov. The dataset itself is still available, as it was originally linked as an Excel spreadsheet on data.gov. As long as the link itself existed, it can be re-downloaded and saved.

However, the fact that it took less than a month for that dataset to be discovered in one repository only for it to disappear, is fascinating. What kind of repository rules exist on data.gov? The dataset itself is for a timeframe from 2000-2013. Could that have a role to play in its removal? Was it simply too old?

This is a fairly important dataset as well, particularly for the citizens and visitors to the state of California. Certainly, violent crime is ever-changing, sometimes for the better, sometimes for worse, but there are patterns that are worth tracking even 10-20+ years later. If that dataset isn’t going to be available on data.gov, it definitely does need to be available somewhere. The repository offered in the repository profile, the National Archive of Criminal Justice Data, would be an adequate fit for this data, but it likely isn’t the only repository available. In fact, it’s certainly arguable that this is the type of data that would be useful in many different repositories that monitor and track this particular brand of data and metadata.

Indeed, it’s quite obvious that this data was meant to be as accessible as it possibly can be. It was, after all, created within a spreadsheet. At least of this is probably because it was taken into consideration that Excel is widely available, fairly intuitive, and otherwise easy to use.

**STATEMENT OF COPYRIGHT LICENSE**

One of the reasons this dataset is likely accessible is, again, because of the type of data it is. This is data gathered by the Department of Health in California, then given to the FBI for further tracking. This is, by its nature, public information. Frankly, putting any copyright license on it would defeat its purpose.

**STATEMENT REGARDING HUMAN SUBJECT CONSIDERATIONS**

This dataset is regarding human subjects: those who commit crime and those who are victims of it. However, this is looking at violent crimes not in the micro sense, where one might wonder if there were names available (there shouldn’t be, even then), but in the macro: it is tallying the number of violent crimes in the state of California for a fourteen-year-span after all. Even if the creators of this information were interested in names, they would likely bog down the data and metadata involved.

When viewing this from a purely ethical lens, the logic makes sense as well. Crime victims and the accused have a right to privacy. That is paramount when discussing sensitive data, even when it’s simply being broken down by the numbers as this dataset is.