# Step 1:

**Task:** Match Table of Contents From Ordinance Aggregators to Census of Governments Dataset

**Code:** merge\_toc\_with\_cog.py

**Description:** We scrape the table of contents from each aggregator (Ordinance.com, American Legal Publishing, and Municode) and then merge the data into the Census of Governments dataset. We use data on the name, county, state, zip code, address, and website of a muni to help merge. We cannot uniquely match all munis with this limited information so we manually match about 2.5% of munis/townships. We do this step first to drop non-muni/township documents (for example county documents) before the actual scraping process.

**Output:** An Excel called matched\_munis which contains a row for each muni/township from each source along with its corresponding data in the Census of Governments dataset.

**Todo:** Clean this code up

# Step 2:

**Task:** Web scrape each municipality

**Code:** American\_legal\_v2.py, municode\_v2.py, and ordDotCom\_v2.py

**Description:** For each source we loop over matched munis for that source and attempt to scrape the text. Sometimes, a muni/township is presented in the table of contents, but, the ordinance is still not available. Thus, we lose some coverage in this step.

**Output:** A pickle file for each municipality containing a hierarchical dictionary that represents the ordinance. Each file is named as “Muni Name”#”PID6 Unique ID”#”State”.pkl.

# Step 3:

**Task:** Filter out non-zoning documents and consolidate muni/township coverage

**Code:** filter\_scraped\_text.py

**Description:** For each row in matched\_munis we check if the ordinance was successfully scraped (does the file exist and is it at least .05 megabyte). Then, we check whether the document includes common phrases included in zoning documents. If it does not, we assume the document is not about zoning at all and drop it from the analysis. We then set a priority to use ordinances from Ordinance.com, then Municode, and then American Legal Publishing when we have multiple sources for the same muni/township.

**Output:** An excel called ‘Sample Data’ which is a subset of the census of governments dataset. We also include the source for each muni/township to help query the ordinance in our database.