The files in this repository are the result of web-scraping cases with docket numbers beginning with “LT-“ from fjdclaims.phila.gov.

The format of the .txt files is comma-delimited with header rows and double quotation marks. The Excel spreadsheet “column-definitions.xlsx” contains an explanation of what each column in each file means.

If you want to work with raw addresses of tenants, look at property-addresses.txt and party-names-addresses.txt.  The docket pages have a box called “Additional Information” containing a “Property Address.”  Usually the address is the same as the address of Defendant #1, but sometimes it is missing and you have to look at the address of the defendant and assume that it is the same as the address of property in question.  The “Property Address” can be found in “property-addresses.txt” and the names and addresses of parties are in “party-names-addresses.txt.”

I have attempted to clean up the addresses as best as I could in the file tenant-addresses-cleaned.txt.  One of the problems with the older addresses was that the parts of the address were squished together into an address like “02218 WCOLUMBIAAVE” instead of “2218 West Columbia Avenue.”  Using a data set of all the properties in Philadelphia (<https://www.opendataphilly.org/dataset/opa-property-assessments>), I built a key to unpack these scrunched addresses.  I also tried to adjust for common misspellings and extraneous/incorrect street directions.  I then ran the addresses through the Google Maps Geocoding API to normalize the addresses.  I was able to obtain a recognizable address in 99% of cases, but I would not be surprised if there are mistakes in the addresses that I determined.  Overall, the accuracy of my clean-up methodology seems pretty good, but there might be particular streets that the system gets wrong. I am sure there are some problems with streets whose name changed between the time of the case and today, and I would not be surprised if some of the addresses in the raw data refer to streets that never existed.

Note that if there is a counterclaim filed in the case, the landlord shows up both as “Plaintiff” and as “Defendant #2.”  So when I made tenant-addresses-cleaned.txt, I either relied on the “Property Address” or the address of “Defendant #1.”

The file “summary-table.txt” contains an analysis of each case on the basis not only of docket entries (which are in “docket-entries.txt” if you want to analyze them) but also the PDFs of the complaints (which I can give you if you want – they are only available for mid-2005 to the present).  Some of the fields are only available for particular periods.  Some fields come from reading the complaints, which are consistently present only from 2006 onward.  Also, in 2011 there was a change in the formatting of the complaints, so some columns have information that is only available for 2005-2011 or 2011-2016.

The cases where “Alias Writ Served” is true are the cases that went the furthest in the process; the service of the alias writ is a prerequisite to execution of the writ of possession.

The data for 1969-2015 were downloaded in late June/early July 2016. The data for 2016 were downloaded in February 2017. Keep in mind that for the 2016 cases, data about the results of cases may be missing because the necessary events in the case have not happened yet.

Let me know if you have any questions.  I tried to be brief in the “column-definitions.xlsx,” but I could provide more precise details if you need them.  I can also re-write the code that generates “summary-table.txt” if there are particular issues you would like me to analyze. If you would like to see the Perl source code that scraped the data, I can provide that.

--Jonathan Pyle, jpyle@philalegal.org