

Recently Released Public Data Sets

ProPublica announced the [ProPublica Data Store](#). ProPublica is making available the data sets that they have used to power their analysis. The data sets are bucketed into Premium, FOIA data and external data sets. Premium data sets have been cleaned up, categorized and enhanced with data from other sources. FOIA data is raw data from FOIA requests and external data sets are links to non-ProPublica data sets.

The costs range from free for FOIA data to a range of pricing (\$200-\$10000) from premium data depending on whether you are a journalist or academic and the data set itself. While the FOIA data may be free, ProPublica has done a significant amount of work to bring value to the data sets. For example, the **Medicare Part D Prescribing Data 2011** FOIA data is free but does not contain details on DEA numbers, classification of drugs by category and several calculations. You can download an [example of the data](#) to see the additional attributes surrounding the data.

The Centers for Medicare & Medicaid Services recently made Physician Referral Patterns for 2009, 2010 and 2011 [available](#). The physician referral data was initially provided as a response to a Freedom of Information Act (FOIA) request. These files represent 3 years of data showing the number of referrals from one physician to another. For more details about the file contents, please see the Technical Requirements (<http://downloads.cms.gov/FOIA/Data-Request-Physician-Referrals-Technical-Requirements.pdf>) document posted along with the datasets. The 2009 physician referral patterns were the basis behind the initial [DocGraph](#) analysis.

Over the next few weeks, I'll be diving in to this data and writing about my results.