

Insider Trader and Publicly Available Criminal Information


Nina Kuklisova & Guillermo Monge

So far in our class, we've seen many examples in which there is a clear necessity to protect certain kinds of information about individuals. However, it can be argued that some data needs to be publicly available to protect or enhance public safety. As an example of this type of possibly "publicly-necessary" information we studied the availability of different criminal listings and court records in the United States. it is important to be able to track people's criminal record.


Our investigation lead us to several background check and official county and state-level court databases (For a extensive but non-comprehensive list of these sites see references [1] and [2]).

Throughout our investigation two major issues caught our attention: the information's availability, usage and limitations vary on several dimensions (the two most apparent: state and type of crime); and the lack of specificity of the information was surprising.

The information's inconsistent usability ranges from states which have user-friendly portals in which the user could search by first and last name, as well as add some additional filtering (action, dates, etc...) to opaque systems which can only be queried by the assigned Court Case Number, making futile any attempt to locate any information from a specific individual. Worth mentioning as well, are the different policies or terms of service agreements: most of these coincide on disallowing any type of automated querying or scraping (some even enforce human use through CAPTCHA) and all warn on the possible inaccuracies of the data.



Aiken County
Second Judicial Circuit
Public Index



[Aiken County Home Page](#) [South Carolina Judicial Department Home Page](#) [SC.GOV Home Page](#)

Neither the County nor any agency, officer, elected official or employee of the County, warrants the accuracy, reliability or timeliness of any information on this web site and shall not be liable for any losses caused by such reliance on the accuracy, reliability or timeliness of such information, including, but not limited to incidental and consequential damages. This publication is provided "as is" without warranty of kind, either expressed or implied, including, but not limited to, the implied warranties of merchantability, fitness for a particular purpose or non-infringement. The County maintains all immunities as provided in the South Carolina Tort Claims Act.

Portions of such information may be incorrect or not current. Any person or entity who relies on any information obtained from this web site does so at his or her own risk. In addition, nothing contained within this web site is an official record of the County or the elected officials responsible therefore. All official records of the County and the offices of countywide elected officials are on file in their respective offices and may be reviewed by the public at those offices.

A person preparing or filing a document for recordation or filing in the official records may not include a social security, driver's license, state identification, passport, checking account, savings account, credit card, or debit card number, or personal identification (PIN) code, or passwords in the document, unless expressly required by law. An individual has a right to request a register of deeds or clerk of court to remove, from an image or copy of an official record placed on a publicly available Internet web site or on a publicly available Internet web site used by a register of deeds or clerk of court to display public records, any social security, driver's license, state identification, passport, checking account, savings account, credit card, or debit card number, or personal identification (PIN) code, or passwords contained in an official record. The request must be made in writing and delivered by mail, facsimile, or electronic transmission or in person, to the register of deeds or clerk of court. The request must specify the identification page number that contains the social security, driver's license, state identification, passport, checking account, savings account, credit card, debit card number, or personal identification (PIN) code, or passwords to be redacted. There is no fee for the redaction pursuant to request.

Access to the South Carolina Judicial Department Public Index web sites by a site data scraper or any similar software intended to discover and extract data from a website through automated, repetitive querying for the purpose of collecting such data is expressly prohibited.

I have read and accept this disclaimer and responsibility for the use of this information:

CRSWeb 6.3 © 2013 South Carolina Judicial Department • All rights reserved

Dataset's inconsistency of use is not only dependent on geography, but on the type of crime as well. Our investigation started with a consistent search of any publicly available data on Insider Trading cases occurred in the US. The Securities and Exchange Commission (SEC) maintains an online dataset [3] of Insider Trading Form 4 Reports. However this dataset has only a small window of time available for free. Further queries, which date earlier than 120 days, and have less or no delay at all are available as part of the monthly subscription rates [4].

The second issue relies on the specificity of this information, the degree by which the information provided in the dataset helps identify the specific individual from the group of individuals that share common traits (same first and last name, same county). The implications of the lack of specificity are probably more apparent on the infamous "no-fly" list [5] which bans the individuals with specific names from boarding any plane. The secrecy of this list has led to many cases in which persons with the same name as known or suspected terrorists have not been permitted to fly. The use of background checks for job position candidates, bank accounts, or other similar services usually rely on the possible criminal records of individuals, therefore false positives may costly and unfairly impair an individual's opportunities, and are sometimes difficult to disprove or have irreparable impact.

Our project on the SEC's dataset of Insider Traders, and the possible impacts of this dataset to the individuals identified in it, in comparison with the impacts of court records or similar public documents.

Our goal was to be able to exploit this dataset to enrich other datasets, therefore tainting the insider trader's reputation or flagging as inappropriate for some activities. We performed multiple cross reference and string comparisons with the names and companies provided for different political campaign contribution datasets (NY State Campaign [6] , and Federal 2015-2016 Campaign Contributions [7]). Our initial hypothesis was that we might be able to identify some overly suspicious generous donations somewhat aligned possible Insider Trading allegations. However, the scarcity of the Insider Trading data and time window limitations of the public free dataset did not allow us to correctly match both datasets.

As a second approach, we performed several string parsing activities to obtain the first name and last name of the Insider Relationship field of the table. A sufficiently adequate algorithm was able to extract a vast amount of first and last names correctly. These names were used to search LinkedIn, and in conjunction with the company involved in the Insider Trading file, some matches were obtained. This slightly fruitful exercise illustrates the possibility of, with appropriate time, and a more

complete Insider Trader dataset, an Insider-Trader-check webapp could be implemented, having clear reputational risk on

these individuals (providing a more similar experience to their less economically endowed criminal counterparts).

This investigation also raised a third potential issue that has been previously discussed in class: the right to forget, especially in relation to criminal records and their right for reinsertion into society. However, we do not want to establish a guideline to an already slippery issue, rather we want to point out and emphasize that this right (or lack of) should be made consistent (or discussed to be) throughout different types of delinquency.

This call for consistency is one of our main conclusions. We find unfair and unjust, in similar ways as discussed in the Belmont Report [8], the different effects of these public datasets, having lasting effects on both criminals and non-criminals. Furthermore, we think it is a matter of grave importance that these biases be considered and, as stated in the Ethics Framework for a Learning Healthcare System [9], an active approach to their remediation should be taken and prioritized.

We believe the problem of specificity is better addressed by expanding the information on individuals from these public datasets. By using other quasi-identifiers, such as education, physical traits, previous job experience, etc... we can reduce the number of “hits” for each name in the database, increasing overall accuracy by reducing false positives.

As a final note, we find this conclusion to be an interesting and novel approach in which the privacy of the greater public is defended and maintained through the publication of more information of specific individuals, rather than by the restriction, anonymization or aggregation of the dataset.

REFERENCES

[1] **Court Records:** [South Carolina](#) (search by name), [Indiana](#) (only search for Case #) , [Iowa](#) (CAPTCHA, search names, pay for advanced), [Oklahoma](#) (search by name), [Arizona](#) (CAPTCHA, search by name), California: [list by county](#), [not publicly available](#) (for fee)

[2] Webpage for performing background checks. This page accesses and queries many of the court records, most wanted criminal listings and other official sources for obtaining all possible criminal records associated to a specific person. [link](#)

[3] SEC Insider Trading Reports: [link](#)

[4] SEC Insider Trading Report Subscription pricing: [link](#)

[5] No Fly List: [wiki](#), [official TSA list](#)

[6] NY State Campaign Contribution dataset: [link](#)

[7] Federal Election Commission - Campaign Finance RPTS and data : [link](#)

[8] *The Belmont Report* - [link](#)

[9] *An Ethics Framework for a Learning Health Care System: A Departure from Traditional Research Ethics and Clinical Ethics* - [abstract](#)