Turning Data Into Clients
What To Do With Golden Arrow Daily Reports[1]

INTRODUCTION

Law firms everywhere are awash in data. Reports come in daily listing new proceedings from administrative courts, new lawsuits filed in state and federal judicial courts, new events that might lead to litigation, or new business transactions. Many of these reports are provided in both a traditional newsletter format and as a data file that could be imported into a database and analyzed.

But how does a law firm filter that flood of data and turn it into actionable information? How do you pick out the leads? It is not realistic to expect an attorney to read hundreds of lawsuit summaries looking for the three or four that actually involve a current client, or a company the attorney would like to target for new business. In our experience, there are four methods of tackling this problem, all with pros and cons.

The purpose of this white paper is to describe each of the four. We draw on our many years of experience working with data in a large law firm, and hope this analysis will help you maximize the value of our products.

THE SCALE OF THE PROBLEM

During one full week in September 2018, Courthouse News reported over 20,000 lawsuit filings in the United States. Golden Arrow's own alerts (covering a half dozen state and federal agencies) average between 700 and 750 regulatory filings each week.

For each of these filings there is at least one party listed, but most filings will list a number of "defendants". It would not be a stretch to expect to see well over one hundred thousand individuals and companies finding themselves on the receiving end of a lawsuit or regulatory agency action each and every week - over five million a year!

As if the volume of data weren't daunting enough, there is the ever present problem of bad data. Plaintiffs do their best, but often get business names wrong. For example, perhaps the plaintiff should have sued the owners of the convenience store down the street, or the corporation that operates the franchise in the region, but they got it wrong and are suing the holding company that actually has nothing to do with whatever tort has befallen them. On the other side, a firm's internal client list may not be the cleanest source of data - errors and limitations that have never been corrected result in a client listed as "[Large Bank], A New York Corporat" (sic) in a firm's internal systems. The data is good enough so the bills go out and the firm gets paid, but for computerized matching purposes, it is problematic.

Law firms and attorneys wishing to tap into this volume of potential clients must develop a strategy to manage this volume of data that flows into their operation each and every day. This suggests that some data management strategies may work better than others. It is also important to remember that in law firms, partners with billing rates over $1,000 per hour are probably not in a good position to be spending time on processing the high volume of potential clients each day. In other words, law firms and attorneys must have a better system that works for them.

[1]And similar reports from other sources.

MANUAL REVIEW

The most obvious method for working through this kind of data is manually: have a person or persons read each report looking for new events involving clients or target businesses. However, manual review will be limited by people's ability to memorize client names. Anthropologist Robin Dunbar suggests that the average person can remember up to 150 individuals as a matter of memory capacity.[2] Yanjaa Wintersoul of Sweden is a world record memory competitor, who in 2017 was able to memorize 212 names and faces in 15 minutes.[3] Realistically, manual review will require some kind of database.

The workflow would involve the person looking up each targeted party listed in a report in a database, then getting the name of the attorney or attorneys who are interested in reports on that party, and then sending an email to the attorney or attorneys. If the person is also tracking non-client businesses, there will likely be a separate database with those businesses and interested attorneys, which will only slow down the process. Fortunately, this is a process that is easy to speed up - if you want to review more reports in less time, hire another person.

Anecdotally, we believe a single person could thoroughly review between 200 and 400 new events per day, depending on the number of clients being searched, the number of parties in each event, etc. This number can be increased if the person is some kind of magician who has memorized all your clients, but as noted above - the limit for being able to remember and recognize individuals has a cap that may not be sufficient in today's legal business environment. For a large law firm with several thousand clients with active matters (and many more that are inactive), this is either going to be slow, require many staff and be expensive, or it will mean the number of parties being searched for is relatively small.

MANUAL MONITORING AUTOMATED BY A THIRD PARTY

Most companies that provide the types of data discussed in this paper offer an option to input information to be searched against the data they deliver. The results of those searches are then delivered to designated recipients, usually via email. This feature is often known as a "Dinger", an "alert", or a "tracker". This is a simple method to increase the number of matches made between the law firms' list of interest and to improve upon the time constraints inherent in the basic manual review.

This method requires the subscribing law firm to rely upon the data matching algorithm employed by the third party service that offers the information, or for the person assigned to create the running search to be skilled in formulating a successful Boolean query. Surprisingly, the level of accuracy differs from service to service and it is therefore important to monitor this aspect of the commercial third party services. Expect frequent false positives while you go through the process of refining your search strategy. Conversely, there will likely also be a large number of missed lawsuits as clients with different name variants are worked into your searches.

It is important to note that using the matching/alerting feature provided by the third party services does require a significant commitment of staff time. The planning stage involves decisions as to what parties to monitor; active clients only, inactive clients, potential clients, types of cases, etc., ... all must be weighed against each other in a cost benefit analysis. Will one set of matches be more productive than others and will adding more justify the time spent in setting these up? Are some jurisdictions more important than others? Are other limitations going to be incorporated into these searches? Then there is the reality that some attorneys will request additional customizations not required by others, and of course some desired customizations simply can't be accomodated.

This process could be completely decentralized and each attorney becomes responsible for their own business development strategy, though our experience suggests that this approach will yield uneven results: some attorneys will have both the skill and time to input and manage their own searches, and others will lack one or both. Duplication is a concern with this approach as well. Not only is it a waste of time but it increases the risk of multiple attorneys pitching work to the same client, never a good outcome.

For these reasons, we believe it is better for a single person or department of trained researchers be designated to manage how the information is input and maintained. A smaller team of expert will be more effective and probably give better results.

Our experience also suggests that this kind of monitoring is almost as labor intensive as full manual review. Stringing together successful Boolean queries is an art and can be hard to do. Alerts need to be kept up to date with changes in the list of clients. Attorneys arrive and depart with regularity. Some services charge extra for these custom searches to be run daily, so that added cost will be a factor. Lastly, not all services offer this feature. It may work well for monitoring lawsuits filed in federal district courts, but you will probably miss new unfair labor practices complaints filed with regulatory agencies like the NLRB.

[2] "Social network size in humans" Hill, R A; Dunbar, R I; M. Human Nature : An Interdisciplinary Biosocial Perspective; New York Vol. 14, Iss. 1, (2003): 53-72. Notes that earlier studies posited numbers between 250 and 5,000 but Dunbar settles on 150 using more recent methods to measure.

[3] "Yanjaa", https://en.wikipedia.org/wiki/Yanjaa

OUTSIDE PARTY HOSTING AND MONITORING

Some services now offer a partially, automated system. Rather than having skilled staff create Boolean queries, the service will do that work for you. All your firm needs to do is upload a list of parties to monitor and the addresses to which the alerts are to be sent.

On the plus side, this will require less ongoing work: once the system is in place to create the necessary data file and upload it to the service, it will probably no longer require any staff involvement, except when things inevitably go wrong. Services like Courtlink have a lot of experience in this kind of project, so the quality of matches is likely to be fairly high - fewer false positives. And, of course, when something goes wrong there will be other people working to figure out what's wrong and make things work.

Commercial services that offer this level of hosting will likely have had experience doing so in a number of law firm environments and bring a high level of success and sophistication to the process. They may also have strategic partnerships with other legal tech providers and can offer additional related products as a result.

On the down side, most law firms will probably not be happy making a copy of their client list and uploading it to a third party, no matter what security promises have been made by the third party.

Another drawback to the use of a third party provider is that a law firm may be limited to using only the content provided by the outside host. That may be acceptable is the attorneys are only interested in one type of lead alert but there are others that may also be desirable. However, if a firm wishes to combine litigation, regulatory, and corporate data into an automated system they will need to coordinate the technical requirements with each provider in order to be certain that the data has usable structure. Some providers may be reluctant to participate.

Also, an outside hosted system will probably require a law firm to make multiple technical commitments. Building a system to generate the specially formatted list of clients often requires programming, and the law firm may want to do more than just have the third party send emails directly to the attorneys.

INTERNALLY DEVELOPED SYSTEMS

If you have programming staff with the skills to implement it, building an system to monitor data can be ideal. As a programming problem matching a list of litigants against a list of clients is not particularly difficult, even when the plaintiffs' names are not always spelled correctly. This is an opportunity for a system to incorporate machine learning and artificial intelligence (there, we said it).. A sufficiently skilled developer could build a system that learned from false positives to improve accuracy. But you will still need a human being to review search results and identify false positives. Like any system that uses machine learning, someone will need to train system.

An internal system could also be flexible enough to incorporate data from multiple sources. In addition to NLRB reports from Golden Arrow, you could incorporate new lawsuit data from vendors like Courthouse News, or business transactions reported by Dow Jones. Once the hard part of matching client names to litigants has been done, feeding in different lists of events from a variety of sources is a straightforward programming task. Another advantage of an internally developed system is you could have different kinds of monitoring. For example, perhaps you are only interested in lawsuits involving clients in a particular region; or only interested in certain kinds of lawsuits. A truly flexible system could incorporate both monitoring for clients and more complicated boolean tests.

The downside of an internal system, even one with artificial intelligence, is that it will require staff to monitor and train. In our experience the best you can hope for is to automate a large part of the manual review process and make it possible to handle a much larger volume of data. A well designed system can take thousands of new events and filter out those that are obviously not of interest. It could also highlight the parties that are potentially a match to clients or targets. One person could easily review over a thousand new matters a day.

CONCLUSION

In the best of all possible worlds turning data from Golden Arrow into new business could be completely and reliably automated for pennies with guarantees of no false positives and no missed opportunities. If plaintiffs' attorneys always used the exactly correct name for businesses on their pleadings; if your client list was good, instead of good enough. If your client list contained not just the proper names of the businesses you represent, but the alternate forms that a human would recognize instantly but a program won't be able to match out of the box.

However, we live in the real world. No matter how you slice it, turning new event data into business is going to have tradeoffs. You can set people at the problem and hope that the business generated is sufficient to offset their salaries. You can use more intelligent systems hosted by third parties, but still need staff to set up the system and maintain it. You can outsource the process, but have to accept that someone outside your firm is going to have your list of clients or targets and system changes are made on someone else's schedule. Lastly, you can implement something yourself, if you have the developer and infrastructure resources to make such a thing happen. You could even use a hybrid of all four approaches, using the method or technique that best suits your environment. For example, you might rely on an outside vendor for a small subset of your client list, but monitor regulatory filings via an internal system. All approaches have trade offs, and it will be up to your own analysis to decide which solution works best for you.

In the end, the first part is identifying the match between a client or target and their need for representation. The next part is what to do once the match has been made - and for that you'll want to read our next white paper.