

Data Analytics: How Parties Are Using Tools Beyond TAR

Law360, New York (November 10, 2016, 12:03 PM EST) --

Google Inc., Amazon.com Inc., Oracle Corp., SAP AG, Teradata Corp., Dell Inc., HP Inc., IBM Corp., the list goes on. What do all of these companies have in common? Big data analytics. All of these companies are leveraging analytics to try to predict the future. What will you buy, where will you shop, how can they advertise to you to influence your decisions and determine what will be trending in the future so they can make investments today. Companies are increasingly using powerful tools to make strategic decisions about where to focus their limited resources and efforts.

When we talk about the same analytics in the e-discovery world, we tend to focus on a few well-known subject areas: technology-assisted review, predictive coding, concept clustering and general review-centric technologies. Although those applications are quite useful in the era of big data, it is important not to forget that there are other uses for analytics that can help parties prove their claims or defenses. Using visualization tools that help judge and jury to see data, and big data reporting tools to present data from differing viewpoints, we can get a strategic view from an aggregate level that oftentimes reveals trends, gaps or other behavior that can be missed when analyzing information on a document-by-document, or record-by-record basis.

These tools provide the ability to zoom out to view a case on a large, year-by-year scale, showing sweeping trends and patterns, or to zoom in to see the same activity on a minute-by-minute scale, or even second by second. In the short term this data and these tools allow parties to more effectively argue their cases, in the long term these tools may very well begin to remove questions of fact from the litigation process by providing definitive evidence of what happened and when. It will no longer be a question of what an individual did or did not do, the data and the analytics will clearly answer that question. Instead, it will become more and more a question of what does the law say rather than what are the facts.

Below are a few examples of how analytics removed or minimized the question of fact from litigation and how parties have leveraged these tools to argue their cases.

Case Study 1: Product Liability

Large-scale data analytics told a great story regarding the purchase of an allegedly defective product. The plaintiff alleged that the defendant sold a defective product, and that this product caused great



Andrea D'Ambra



Charlie Platt

personal harm to the plaintiff's family. The plaintiff lacked documentation of the sale, instead relying solely on sworn eye witness accounts of the sale.

Using large-scale analytics on data related to individual store sales, shipping logs, manufacturer purchase orders and store returns, the defendant developed and argued a strong case that it had not sold the specific product in question to the plaintiff in the timeframes alleged.

The defendant did identify sales of the specific product, but all sales preceded the alleged sale by many years. In fact, a very clear pattern emerged from the compiled data — the product had been replaced with a newer model roughly two years prior to the date in the plaintiff's complaint.

With access to the raw data from multiple systems, and the ability to run unified analysis, the defendant was able to remove the question of fact, did they sell this product to the plaintiff, and effectively proved that the plaintiff's case lacked merit. Once it became evident that the plaintiff did not purchase the specific product listed in the complaint from the defendant, the case settled quickly in the defendant's favor.

Case Study 2: Insurance Fraud Class Action

In some cases access to data is only the beginning. One recent case included dozens of systems and multiple data transfers between systems. The plaintiffs alleged large-scale, systemic insurance fraud, and had filed for class certification. The defendant was caught off guard at the outset of litigation, and lacked the information required to know if the plaintiffs' had a good argument or not. In developing their defense, the defendant analyzed large quantities of data, and was fairly certain they were not unjustly enriching themselves. Nevertheless, initially the defendant was unable to clearly explain how the process worked and where the money eventually landed. They needed to understand how the data flowed between their various systems, both inside and outside of the company.

No single employee knew how all the data fit together. Individuals responsible for a specific business function knew their system, but didn't know what happened to the data when it left their area. Through interviews and testing, the defendant was able to compile a map of the overall system architecture and data flow.

After building a model to show how data flowed through the organization, the defendant was able to follow the money through the various business processes. With this model, combined with the analysis of the data, the defendant educated the plaintiffs on how insurance payments flowed through the organization, and, again, using analytics to remove questions of fact, demonstrated that the plaintiffs' case lacked merit. The result? The plaintiffs voluntarily withdrew their complaint.

Case Study 3: Quickly Gain Insight Into Opponent's Productions

Leveraging data analytics to assess the completeness of an opponent's production and to prove spoliation can be very effective and helpful. Using visualization software we can show trends and correlations among key custodians or between a party's production and its opponent's production, highlighting areas for concern and identifying documents for further research.

In one case, the plaintiff alleged the defendant violated their nonsolicitation clause in their contract. The defendant argued that the plaintiff's CEO had sent emails and text messages outlining an understanding between the parties that permitted the hiring of certain employees. Unfortunately, the plaintiff's

production of emails and text messages was suspiciously small. The defendant charted email volumes over time from metadata contained in communications to highlight gaps and dips in quantities of email over the timeline of events. By quickly visualizing the unstructured content, the defendant found not only dips in communication patterns with key individuals from the plaintiff's email, but also identified vast differences in email message counts when compared to communications the defendant had collected and produced. Emails that the defendant had produced that should have appeared in the plaintiff's production were suspiciously absent.

Applying similar techniques to text message communications, the defendant also proved that key custodians in the plaintiff's organization had intentionally and selectively deleted text messages during the most critical period in the case. This ability to show targeted deletions disproved the plaintiff's position that the custodians were in the practice of mass deleting texts as they were received. At the hearing, exhibits visually demonstrating these facts helped prove and win a spoliation motion in favor of the defendant.

Case Study 4: Wage and Hour Class Action

In most cases involving electronically stored information, user-created files (i.e. electronic files that are created by users vs a device's operating system) are the most common file types analyzed and produced in litigation. These are what we generally think of when we think of ESI, for example email, PowerPoint, Excel, PDF and MS Word documents.

We recently worked on a matter involving a claim for unpaid overtime. The defendant used analytical tools to stitch together forensic artifacts found in the employees' computers. These artifacts included data such as browser history and online user activity, more arcane forensic data such as link files and registry artifacts, complex structured data found in files such as Excel worksheets, combined with traditional ESI, such as email and text messages, to illustrate the employee's actual work habits. The defendant analyzed two specific data sources — the plaintiff's laptop, where "all work was performed," and the former employee's timecards. Additionally, the defendant was able to retrieve an iPhone backup file from the plaintiff's laptop, created only days before termination.

Using industry standard e-discovery and forensics tools, the defendant extracted and normalized the metadata and then overlaid the various data sources, piecing together and aggregating the user activity over a two-year time period. This analysis provided a strong argument to disprove much of the time the plaintiffs alleged they had worked outside of their standard eight-hour day and ultimately led to a favorable class settlement for the defendant.

In Conclusion

As evidenced by these case studies, analytics is a critical tool for understanding and evaluating the massive amounts of data being produced in litigation today. Implementing these tools into the litigation process provides valuable insight into the story behind the data. Incorporating big data visualization into the process allows us to illustrate complex data systems and analysis through simple and compelling demonstrative exhibits.

Ten years ago much of this data either didn't exist or was exceptionally expensive to obtain and analyze. Today there is more data available, it's much easier to obtain, and the analysis, while difficult and requiring expert assistance, can be exceptionally compelling to a judge or jury. As we move into the future, and as companies continue to amass more and more data, and that data increasingly becomes relevant to litigation, analytics will be a crucial part of the process for parties seeking hard proof to win

their cases. Think of the transformation that electronic data (such as email) has had on cases over the last 10 to 15 years, and realize we are standing on the same precipice today with regard to big data and analytics, only with exponentially more data.

—By Andrea D’Ambra and Sam Sessler, Norton Rose Fulbright US LLP, and Charlie Platt, iDiscovery Solutions Inc.

Andrea D’Ambra is a senior counsel in Norton Rose’s New York office. Sam Sessler is a practice support manager in the firm’s Dallas office. Charlie Platt is a senior managing consultant at iDiscovery Solutions in Washington, D.C.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of the firm, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

All Content © 2003-2016, Portfolio Media, Inc.