Get Smart

FORENSICALLY SPEAKING … FOR THE LITIGATOR

SRR
STOUT | RISIUS | ROSS

How to Deal With a lot of Data (aka – "Big Data")

AICPA
American Institute of CPAs®

"Big Data" listed as the top issue facing forensic and valuation professionals during the next two to five years.

AICPA survey
July 2014

Industry experts say that **"Big Data"** is changing society and that its use within law firms has already exponentially increased. There are multitudes of seminars, webinars and articles available to lawyers as to how Big Data can be used for jury research, case strategy and early case assessment, case mediation and/or settlement, docket management and on compliance matters regarding privacy and data collection and storage. In an article titled "Know Your Judge…Through Analytics, Not Anecdotes" written for the ABA Journal's Persuasive Litigator website, internet blogger Dr. Ken Broda-Bahm provides a number of excellent examples of how Big Data is impacting the legal industry.

Litigator David J. Walton in the article "How lawyers and law firms operate in a Big Data world" from InsideCounsel Magazine, states that *"Big Data has enormous implications for corporate clients … it has important consequences for the practice of law, meaning how attorneys perform legal work on behalf of clients"…"to engage in trial work lawyers must understand the nature of electronic data - where it exists and how to obtain it, authenticate it and get it admitted … and when to marshal Big Data analytics to build a case."*

An unfortunate fact about Big Data is that it brings about a certain "fear", generally as to the uncertainty of how to obtain it and then how to manage it. One myth currently being perpetuated is that Big Data is only for big companies and only applies to highly complex matters. This is not the case. Big Data should be addressed with an objective to dart the bulls-eye with data that is pertinent to the matter regardless of complexity.

FINANCIAL AND OPERATIONAL DATA THAT IS AT THE CENTER OF NEARLY EVERY INDEPENDENT INVESTIGATION AND COMMERCIAL DISPUTE

> "In God we trust. All others must bring data."

W. Edwards Deming,
statistician, professor and author

> "Torture the data, and it will confess to anything."

Ronald Coase,
economist and Nobel Prize laureate

> "Big Data is not about the data."

Gary King,
quantitative methodologist at Harvard University, noting that, while data is plentiful and relatively easy to collect, the real value is in the analytics.

> "If your experiment needs statistics, you ought to have done a better experiment."

Ernest Rutherford, physicist and Nobel Prize in chemistry

> "It's not about Big Data — It's about Good Data."

Datawatch / IBM

Big Data is often associated with so-called unstructured data, or data taking the form of e-mails, text messages, audio or video files and includes information available on social media such as Facebook, Twitter or Linked-In. But Big Data also exists within **structured data, as this is where you generally find the numbers that relate to the relevant financial and operational data pertinent to the matter**. This type of data can be comprised of voluminous detailed transactions such as purchase orders, invoices, checks and wire transfers, journal entries and stock prices and trades.

But at the core Big Data is just that — data. It is essential to know what you plan to do with it, what questions it will answer and where it fits into your case. Or at least what potential benefits it could provide and if it can be mined to provide metrics that might help support your client's case. In the InsideCounsel Magazine article previously noted, Walton provides some thought-provoking words to these benefits when he asserts *"Big Data is already influencing the kinds of arguments made in class actions and other lawsuits that typically invoke statistical sampling, which has historically been used to extrapolate about both cause and effect. To the extent that Big Data, by making it possible to review the entire data set and not just a sample, trumps probability-based calculations, will statistical sampling still be allowed? The answer likely depends on the specific factual and legal issues in dispute."* While it may be a stretch, Walton's thought provoking assertion that statistical sampling may be rendered obsolete with the advent of Big Data is raising a few eyebrow and perhaps the ire of forensic economists. However, we certainly agree with Walton's prognosis that **"There can be no doubt that the use of Big Data analytics will create many opportunities for lawyers to tell a much richer and more real story than they have in the past."**

Obtaining Big Data and marshaling it to the point of useful analytics is highly correlated to the cooperation, depth of skills and resourcefulness of the IT department. Considerations regarding these matters should be contemplated in parrallel with the exploration of data availability and its potential use. Early in the process, several important considerations and questions surrounding Big Data are: "How much data will there be?", "Where will our legal and forensics team store and process it?", and "Who will perform the tasks of validating, filtering, finding and processing the data you needed for the case?" There are times it may be possible to utilize your client's IT environment and staff to perform part of the initial blocking and tackling steps of data collection and supplying preliminary analytics and metrics. This type of leveraging could result in a significant cost and time savings for the client. However, this is not always possible due to IT resource availability or the nature of the case. In such cases, the IT department will be called on to help identify and target the necessary data, potentially filter it and provide it to the legal team's forensic experts in the most user-friendly fashion. Successful data transfer calls for close cooperaton and teamwork amongst all parties involved.

The next page provides highlights of how Big Data was used to help support and build a case using a real life example. This will hopefully provide you with some insights and help alleviate some fears or myths in helping you **Get Smart** about **Big Data!**

> "We are not enemies, but friends. We must not be enemies."

Abraham Lincoln,
1st inaugural address 1861

Applies to the legal team and IT department relationship.

THE IT DEPARTMENT

Example: Securities Broker-Dealer

An SEC registered securities broker-dealer (defendant) was accused of violating certain securities trading practices and market manipulation over a multi-year period by a group of electronic traders (plaintiffs). From a data perspective, the heart of the matter involved hundreds of thousands of trade orders presented by the plaintiff to the court. The plaintiffs alleged that these were orders not executed by the defendant or orders that were "incorrectly" executed by the defendant. The plaintiff accused the defendant of taking advantage of the advance knowledge of pending orders (i.e. front running), causing economic harm to the plaintiff. One important question to the lawyers and forensic experts working on behalf of the defendant was, ***"Given the large volume of data, how can we mine it to discover new information that is relevant to the case? And what impact does any new information have on the allegations of "harmed orders" against their client?"***

The client was not able to utilize its IT department's environment to perform the anticipated data mining and analytics expected. Data storage and data processing had to be performed off site and in part using an outsourced IT environment. An independent forensic accounting firm was hired by the defendants to serve as expert witness and to also perform the data management and analytics work. This included collecting and hosting over 200 terabytes of data provided by numerous and disparate sources including plaintiff and defendant proprietary systems and market data from five public exchanges. The data spanned for over six years and was provided on various forms of media including: hard drives, tape cartridges, DVDs, and hardcopy. Billions of records (individual transactions containing pertinent data such as quote, order, trade and associated fields/attributes) of "structured data" were imported and converted from a variety of compressed formats and structures including Oracle databases, DB2, CSV, text logs, and other plaintiff proprietary systems. This data was arranged and normalized into a functional and understandable manner for use by the lawyers, which required the data to be filtered, reconciled, and mapped to establish the relevant associations amongst the disparate data. Comparative analytics, sampling, and customized queries (primarily using Microsoft SQL, Access, and Excel) were then performed to identify specific examples and prepare economic models from the data in order to identify dispositive explanations for how the allegedly "harmed trades" were ultimately administered by the defendant.

DATA FINDINGS: There were inconsistencies and anomalies between the plaintiff's data and the market exchange data, which identified holes in the plaintiff's complaint. This information was not previously identified or accounted for by the plaintiff, and tainted plaintiff presented analysis and models. In addition, many anomalies and errors were also uncovered with the population of "harmed orders" presented by the plaintiff such as: potential duplicate orders that were double-counted; orders that were during times of "fast markets" (many of these could not be filled, thus not the fault of the defendant); orders that were executed, but where profits were realized by the plaintiff (benefiting not harming the plaintiff); and orders that were basically invalid with incorrect or incomplete data fields (alpha characters in numeric fields, numeric or special characters in alpha fields, invalid and missing dates, etc.). **This was powerful information for the defendant's lawyers to know. The plaintiff lacked thoroughness, accuracy, completeness and validity with their analysis of the data. There was a lack of quality assurance with the data submitted by the plaintiff to the court.**

Over 50% of the proposed "harmed orders" were eventually associated with some type of data related issue! This case was eventually settled at a fraction of the original amount sought by the plaintiff. **"Big Data"** and forensic accounting played an important role in providing critical information to help arm the defendant to resolve the case and achieve the settlement.

- Remember the basics: validate all data definitions and record layouts. Applications and Systems are continuously being upgraded and can cause changes in core data. Beware!
- "Start with the end in mind." — Stephen R. Covey, *Seven Habits of Highly Effective People*. Demonstrate feasibility and develop prototype using smaller data increments. Create a model using a day or a month of data before attempting to tackle years of Big Data.
- Treat the IT department as a friend, not an enemy. IT professionals can often help with data issues.