

Big Data: Big Pain or Big Gain for Lawyers?

By Sharon D. Nelson, Esq. and John W. Simek
(©2013 Sensei Enterprises)

We are going to assume that readers know nothing about “big data,” including its definition. Though everyone has heard the term thrown around, when we use it in our presentations, audiences look at us blankly. So let us begin at the beginning.

What is Big Data?

Here’s the first problem. There IS no standard definition. Our vastly oversimplified definition is “big data” is so vast an amount of data that it cannot be analyzed by traditional tools. “Predictive analytics,” which uses statistics, modeling, machine learning, data mining and more, is applied to big data in order predict future events, suggest how to manipulate them, identify trends, and much more.

Big Data All Around Us

We are often unaware of “big data” until we hear it referenced. Perhaps you know the TV commercial in which a competitor throws spears at Google by advising viewers that Google goes through all your e-mail (now THAT is big data) and sells the analysis results to advertisers so they may determine which ads will target you. The commercial must be working since we’ve seen it for months everywhere.

The February 25th issue of *NetworkWorld* featured a story called “Big Data Gets Creative at Etsy.” At the website for this company selling handmade crafts and vintage items, Hadoop-based analytics have been used to mine vast quantities of data and turbo-charge sales with the results. Sales jumped 70% last year. Roughly 75GB of customer transaction data is stored each day, aggregated and analyzed. Any of Etsy’s 150-plus engineers can deploy live code to the website at any time – and that happens 20-30 per day based on their data sifting.

If you've ever wondered about the myriad ways in which big data can be used, perhaps by some of your clients, it is well worth reading the article at <http://www.networkworld.com/news/2013/022513-etsy-big-data-266841.html>.

One more story worth mentioning gave us a giggle. The International Association of Fire Chiefs has used predictive analysis of big data to learn that many of the most desirable volunteer firefighters like hunting and country music. As a result, it advises fire departments to hold recruitment events in sporting goods stores and to partner with country music stations to advertise.

Why Should Lawyers Care About Big Data?

Those in the e-discovery world have begun to grasp the implications of big data, but most other lawyers have not. Big data presents both challenges and opportunities for lawyers – so many that it is almost impossible to know where to start. Big data will be used against lawyers and law firms (more about that later) and will become a source of new work and a worry about the implications for work we've always done. Donald Wochna, chief legal officer at Vestige Digital Investigations, was quoted in *Law Technology News*: "Big data in general, and predictive data analytics in particular, are the potential holy grail in the practice of law."

"Fast, high-performing data analytics can help enterprises and law firms harness expanding data collections to guide them on everything from finding profitable efficiencies to making important decisions in case strategy," added Matt Gillis, vice president and managing director for litigation tools and professional services at LexisNexis. "It's not uncommon for attorneys to sort through and make sense of upwards of 300 terabytes of data when preparing for a case [and] the massive volume of data simply outpaces the capabilities of traditional technology tools to process that much information in a timely fashion."

There is so much hype around big data that it requires some tempering. It reminds us very much of the hype around "predictive coding" in e-discovery. More data is not necessarily a good thing. Sometimes, more is simply more – and therefore increasingly complex to manage and secure.

There is no way to cover all the implications of big data in anything less than a book. So we'll try to give you a sampling – a cornucopia of exciting possibilities and unsettling complications. Let us begin with lawyers.

Using Big Data to Evaluate Law Firms

This is the most frightening possibility to law firms – especially large law firms – that clients will begin to use data analytics to evaluate law firms in a far more precise fashion. If, as is so often the case, a large client spreads its work among 10-20 firms, it can aggregate data to determine how much it is being charged by different firms for similar work, what percentage of the work is being done by partners vs. associates, what the hourly rates are and the totals for similar work, what percentage is based on value billing and how much is being billed for travel and other expenses. Comparisons between law firms will be much easier to make as analytic tools are applied. Clients may even analyze the extent to which fees and costs come down after a “Come to Jesus” meeting with law firms – and how long the effects of the meeting appear to last.

Management committees at large firms are going to have a serious case of the willies when such practices become commonplace. Clients will certainly look at measurements of success – percentage of trials won, successful outcomes in arbitration or contract negotiations. They may also evaluate the data and conclude that they can use a boutique firm in North Carolina for far less than a megafirm in New York City.

Clients and lawyers will together use analytics on big data to help predict case outcomes. Think of it as a real world application of the movie **Runaway Jury**, where everything (almost) was known about the jurors. We may learn from large databases of realtime search terms what legal issues are troubling particular communities. Risk management attorneys may be able to figure out where their clients are most likely at risk from studying government data.

Analytics software can speed up management tasks, such as distributing cases, projecting revenues, projecting case budgets, projecting revenues and – most

important of all – predicting outcomes. All of this could help determine that most elusive of client dreams – a fee estimate that makes sense based on previous matters involving similar factors. With all this, it is also possible that law firm management budget forecasts may be more reliable.

Is there an ethics and security impact? Yes, of course. The more data we have in more places, the greater the chance for security holes. Are lawyers required to take reasonable steps to protect client data? Once again, yes. How easy is it to define ‘reasonable’ in the age of big data? That question is much less easy to answer and very few people in the legal field have begun to discuss it, much less reach conclusions.

Big Data in E-Discovery

This was certainly one of the hottest topics at LegalTech in January 2013 – and it was completely absent in January 2012. Why is there so much interest? Perhaps because it could be the harbinger of another e-discovery gold rush. The same new technology which users employ to understand big data can be used for e-discovery of big data – to an industry which has remained married to per gigabyte pricing, the thought of **REALLY** BIG DATA has vendors salivating.

Add to that the fact that big data comes with lots of errors and is messy to corral and vendors are lusting after it. Not many folks in the e-discovery world knew much about Hadoop or other analytic software two years ago, but now everyone seems onboard that train.

A huge headache in e-discovery has been the large amounts of unstructured data, including e-mails, word processing documents, social media communications, etc. constituting a huge volume of data subject to e-discovery. The new predictive analytic tools make help make sense of the tsunami of information, and give attorneys faster, more reliable access to potentially relevant data which needs to be processed and reviewed. The famous (or infamous) predictive coding that has stormed the e-discovery landscape makes use of predictive analytics. As commentators have consistently said, this is an evolving area and we will find

technology-assisted review morphing with technologies and continuing analysis of how to improve the process by experts in this field.

The Eye in the Sky: A Hot New Big Data Topic

We used to think of drones as military apparatuses. No longer. For a measly \$300, you too can be the proud owner of a camera-equipped drone. The higher end drones can cost tens of thousands of dollars and have lots of commercial buyers. Real estate agents use them for aerial photos and videos. Wildlife researchers and search and rescue groups employ them. There are lots of legitimate uses – no doubt of that. But voyeurs may also love “the eye in the sky.” How much privacy you might have in your yard, or in your home, may be changing. It would be illegal under many state laws to hover a drone outside your bedroom window, but then again, who is to say who owns the drone and prove it?

Law enforcement loves drones and is adopting them in record numbers. But turnabout is fair play and the Occupy Wall St. protesters had their own drone (dubbed “The Occucopter”) to monitor the authorities.

Who regulates drones? The Federal Aviation Administration does if they operate at 400 feet or higher. If the drones operate lower, they must follow model aircraft rules. Many of the rules are being put to the test by a drone-happy populace and, yes, the data collected will certainly fall under the heading of big data over time.

Concerns about governmental drones invading privacy are regularly voiced by privacy groups – and in 2013, a bill was introduced in Congress to regulate the use of drones by authorities. Twenty state and local governments were legally operating drones by May of 2012. With the federal government allocating billions for drones, experts expect the state and local drone obsession to intensify as well.

Other Privacy Issues with Big Data

Everyone, by now, has probably heard about the Minneapolis teenage who received all the ads for nursery furniture and maternity clothing after shopping at

Target. Whatever she searched for was duly recorded and used to send the ads. Her parents were more than a little surprised to find out that their little girl was pregnant after they complained about the ads.

Facebook has a huge amount of our data. In fact, it has more than some people think because they never read the Terms of Service which allow Facebook to monitor your online activities while you are logged in. That very valuable data is sold to advertisers so that they can have ads related to your online activities pop up or show on visited web pages. Not so bad if you're searching for a new car, but what if you hang out (while married) in dating sites? Or you frequent pornography sites? What are you doing online that you'd prefer to be kept private? How about searching for help in treating a substance abuse problem? Or how to file a bankruptcy? Perfectly innocent activities certainly deserve to be private but privacy is eroding fast in the big data world.

We just don't think about our digital privacy. We use our car's GPS, we post on Facebook, we buy on Amazon and we use location service apps on our smartphones. We create our own "big data" cloud about who we are, where we are, what we do, what we like and what we don't.

The government took a huge bite out of our privacy at the end of 2012 when it adopted rules which allow the little-known National Counterterrorism Center (NCTC) to examine the government files of U.S. citizens for possible criminal behavior, even if there is no reason to suspect them. In the good old days, the government couldn't store information about ordinary Americans unless they were a terror suspect or related to an investigation.

Now, NCTC can copy entire government databases—flight records, casino-employee lists, the names of Americans hosting foreign-exchange students and many others. It can keep data about innocent U.S. citizens for up to five years (talk about "big data!") and to analyze it for suspicious patterns of behavior. Previously, both were prohibited. Data about Americans "reasonably believed to constitute terrorism information" may be permanently retained.

Online privacy continues to shrink as our data and our activities grow in value. There is fertile ground for privacy lawyers in all of this – and the opportunities will continue to grow.

Compliance and Security

Will we reverse course and, instead of getting rid of all the data we don't need, begin to keep everything? How will compliance laws and regulations handle big data? No one yet knows, because this is a fairly new phenomenon, but there is sure to be more legal work in assisting clients with compliance issues.

Our own view is that keeping everything will never make sense – not in a world of per GB pricing for e-discovery (which will probably be per terabyte pricing in the near future). While it makes sense to keep information that may be useful (or legally required), “taking out the trash” still makes sense to save money in possible litigation. But we will want to keep everything that may be useful in this new big data world for analysis. It is defining “useful” that is causing all the controversy. Balancing compliance factors with big data's benefits and risks will no doubt involve both in-house and outside counsel as data governance policies are developed.

The federal government has begun to warn about the security dangers implicit in big data and security companies have been scrambling to produce big data security products. Once again, clients are going to need advice about this aspect of security and how it impacts risk management – a dicey opinion to give with security management of big data still immature and changing rapidly.

Final Words

We know that we have only scratched the surface of big data. As you have seen from the snippets we've shared, big data will offer lawyers new opportunities for work and for serving clients in a more efficient and predictable way. It will also be used to evaluate the performance of lawyers and law firms, which will be repugnant to many. Avvo and other lawyer rating sites will no doubt get in the big

data game too. No joy in Mudville there. However, early adopter law firms who use big data analysis wisely will be far ahead of their competition. Whatever it brings, big data is here to stay and the task at hand for lawyers is to learn what it is today and what it will be (as our friend Richard Susskind would say) for “Tomorrow’s Lawyers.”

*The authors are the President and Vice President of Sensei Enterprises, Inc., a legal technology, information security and digital forensics firm based in Fairfax, VA. 703-359-0700 (phone)
www.senseient.com*