BIG DATA
FOR LAW FIRMS
We have all experienced the phenomena of frequency illusion: You go to look at a new car and for the first time ever, you see an electric teal-painted Honda Civic. On your way home from the dealership, you pass one turning into your neighborhood, and tonight, you spot one while watching your favorite TV sitcom. Many of us in the legal profession are having that same sensation with big data. Suddenly, big data comes up in conversation – legal business specific or otherwise – and it’s impossible to escape it.

Lately, legal business and technology conferences across the globe are feverishly adding big data to their agendas, and Continuing Legal Education offerings and articles like this one are penned to peak your interest. But what does this mean to you, your role and your firm? Is everybody really doing big data? Is big data the next cloud, maybe even the next iPhone or iPad trend, or just a passing marketing-hyped fad? Can businesses – including law firms – even compete, let alone exist, without a big data strategy?
So, if we buy into some of the hoopla, let’s assume big data will make law firm administrators heroes (or goats?), depending on their firms’ ability and willingness to understand and champion (big) data, define it, scope it, implement it and reap the much sought after return on big data investment. Where will you end up? Let’s start from the top and find out.

**BIG DATA DEFINED, DEMYSTIFIED**

I have often compared big data, especially as it relates to its application and proliferation in the legal marketplace, to *Harry Potter’s* Lord Voldemort: the technology and trend that “must-not-be-named,” simply because of our inherent fear of the unknown and our unpleasant association with it. What is so ghastly about big data anyway? The biggest strike against big data would have to be its overexposure and frequency illusion as mentioned above. If we pull back the covers and educate law firm leaders – the C-Suite, partners, fellow administrators and legal technologists – on the big, big data opportunity, forehead wrinkles will disappear and smiles will reappear.

I have found that defining big data is really a matter of perspective. Technologists have a bits and bytes definition; marketing tends to have a customer profiling and analytics-focused view; governance and IT might frame big data in a security setting; and firm attorneys may see it as their opportunity to counsel clients on issues pertaining to (big) data security, privacy and information management. Generally speaking, the most popular application of big data is with analytics and data mining.

Big data is the collection of data sets so large and complex that it becomes difficult to process using standard databases and data processing tools/techniques. Big data includes structured data and unstructured data – basically anything that does not fit in typical relational databases (such as email, text, video and audio files). In addition, there is the distinction between current data and dark data – information assets that organizations collect, process and store during regular business activities, but generally fail to use for other purposes (for example, analytics or business relationships). Organizations often retain dark data for compliance and storage purposes only, typically incurring more expense and sometimes greater risk than value.

Big data is commonly defined by the **4 Vs**:  
- **Volume**: How much data do you have?  
- **Variety**: Is your data structured, unstructured, text-based, internal and external?  
- **Velocity**: Are you able to analyze data in real time as it happens to enable instantaneous decision-making?  
- **Veracity**: Can the data you are collecting be trusted? Is it accurate?

Beyond determining if there’s even a business case for testing a big data project, the challenges of managing big data include capture, curation, storage, search, collaboration, sharing, analysis and visualization.

**SURVEY DATA ON BIG DATA IN LEGAL**

As you might imagine, there is a dearth of market research and intelligence when it comes to defining big data in the legal profession. Despite an increasing number of Am Law 100 firms forming big data privacy and information governance practice groups, there is a lack of data when it comes to who is doing what, when and how.

In August, the International Legal Technology Association (ILTA), in partnership with InsideLegal, published the 2013 ILTA/InsideLegal Technology Purchasing Survey, which included responses specific to the current and future use of big data among ILTA member law firms. While 39 percent of survey respondents indicated they are either not sure about big data or don’t think it will have much of an impact within the next five years, 15 percent stated the biggest impact will be related to storage capacity. When asked how big data might negatively impact legal IT, 25 percent said it...
will lead to cost increases, and 12 percent said big data will require major user training and result in personnel/expertise drain. Finally, the participating firms were asked about positive big data impact on IT within the next five years. Sixteen percent of respondents stated big data will lead to more strategic use of data via predictive data modeling, data mining and more accessible data analytics.

Several speakers at ILTA’s 2013 Annual Educational Conference underscored the relevance of big data in legal and the vitality of using analytics to better serve internal constituents as well as better compete for existing and new business. During a conference session, “Big Data: The Big Picture (and Big Catalyst?),” Blain Banick, Chief Business Development Officer at Husch Blackwell and session panelist, said, “Big data analytics in particular will save client relationship management in law firms and will be one of the most fundamental changing points in the history of law firm business development.” Justin Ergler, fellow panelist and Director, Alternative Fee Intelligence & Analytics, GlaxoSmithKline, took it a step further: “In order to survive the new marketplace, law firms must differentiate themselves with something other than the excellent lawyering mantra. Leveraging big data is a way to do that.”

I firmly believe the common mental roadblocks keeping us from seeing the bigger picture when it comes to big data in legal are based on the belief that firms do not have enough data to be big, and big data is too new and hyped to have any long term impact.

However, big data use in legal is not new at all. In fact, it has existed in varying forms for at least the past 20 years:

- For decades, firms mined their financial billing data and used business intelligence tools to make sense of this information, speed up invoicing and cash flow, and determine general profitability.

- Since email (dating back to the mid 1990s), data warehousing has been used to store e-files and data. This has taken off since data storage costs have plummeted, new cloud storage options have become available, and search technologies have eased search and find tasks.

- The Federal Rules of Civil Procedure revisions in 2006 made practical changes to discovery rules to make it easier for courts and litigating parties to manage electronic records. As a result, collecting and managing Electronically Stored Information (ESI) has become a big data-esque issue for many law firms and corporations.

**BIG DATA: DID YOU KNOW?**

Putting big data volumes, ROI and general use into perspective

- **Twenty-five percent** of big data use comes from the financial industry: Large, global banks manage 3.8 petabytes (a rough equivalent to 850 million iTunes song downloads) of data on a daily basis, including credit card, mortgage, loan and customer information.

- **Big data analysis of social media** channels using text-mining algorithms can help predict trends in automotive equipment failures; vehicle design and performance; public opinion on sports teams and individual players; and potentially, the outcome of lawsuits and complex litigation.

- **Forty-five percent** of global telecommunications carriers with big data initiatives use location data to support intelligent marketing campaigns, detect fraud and improve network quality.

- **Big data analytics in sports** are used to drive decisions regarding merchandising sales, labor agreements, ticket sales, talent/player acquisition and retention, and TV/digital media deals.

- **Weeks before** there’s a noticeable increase in flu patients coming to local/regional hospital emergency rooms, there is a spike in Google searches for terms, such as “flu symptoms” and “flu treatments.”

- **Zynga**, the San Francisco game maker behind Farmville, processes 1 petabyte of content for players every day, which helped them determine that 1 percent of its players account for up to 50 percent of its revenue. Does your law firm have similar metrics for your clients?

- **The average teenager** sends 4,762 text messages per month.

- **Thirty-two billion searches** were performed last month on Twitter.

- **Currently, there are 2.7 billion Internet users**, which is roughly 38 percent of the world’s population.
Cloud-based file repositories and storage solutions such as Dropbox, Box, Microsoft SkyDrive, Google Drive and NetDocuments are routinely being used by legal professionals to store, manage and collaborate on large volumes of documents and files.

Data mining, analytics, and business intelligence tools and technologies, such as LexisNexis Redwood Analytics, TyMetrix and Sky Analytics, to name a few, are equipped to aggregate and number crunch big data sets.

BIG DATA FINANCIALS, ANALYTICS AND PRIVACY

The legal profession is full of transaction-based processes, so the opportunities to responsibly leverage the data that is a by-product of these transactions, to improve decision making and streamline operations, most definitely exist. InsideLegal has spent at least the last 10 months reaching out to law firms, corporate legal departments and legal technology vendors involved in big data-type projects. Here is a brief sampling of some real world examples:

Legal Spend Management: Corporate law departments can analyze large amounts of data from invoices to determine which factors most influence rates and then negotiate better deals based on that data. TyMetrix Legal Analytics has developed a free mobile application that aggregates data from thousands of law firm invoices and comes up with hourly billing rates based on practice area, firm location, etc. The TyMetrix RateDriver™ mobile application is based on a statistical model published in the Real Rate Report™ – a statistical analysis of legal invoices. Similarly, Boston-based Sky Analytics aggregates a company’s outside legal spend into an interface and instantly provides legal cost control insights. Sky does not require eBilling systems and can process paper or PDF invoices, as well.

Law Firm Helpdesk Metrics: In the user support/help desk arena, law firm support and IT groups can reference metrics (based on the aggregation of millions of help desk tickets) to determine which applications need more support, how much time and resources to allocate for a big technology conversion, which days and hours support will see higher volume than others, and which offices need more training on specific applications. Intelliteach, a law firm-specific outsourced help desk provider, has been publishing these statistics in, “The Guru’s Guide for Legal Service Desk Support: Law Firm Specific Metrics & Key Performance Indicators,” a free industry report, since 2011. The latest version can be accessed here: www.intelliteach.com/wp-content/uploads/2013/01/Guru-Guide-Version-3.pdf.

Predictive Coding in eDiscovery: The most expensive part of the eDiscovery process is undoubtedly document review, with experts citing that nearly 75 percent of the eDiscovery budget goes toward attorneys reviewing documents, often times exceeding millions of documents across multiple jurisdictions. Increasingly, technology is evolving to meet companies’ endless quest to cut costs. Today, technology-assisted review (TAR), also known as predictive coding, uses the expertise of attorneys and machine-learning techniques to automate the prioritization of documents for review, based on how likely they are to be responsive to a particular matter. By using such big data technology better, corporations are reducing costs and more effectively identifying key documents. Also, a number of U.S. courts have endorsed the use of predictive coding technologies.

Big Data Privacy & Information Security: A few U.S. firms have taken the first step and opened separate big data practice groups focused on offering compliance and regulatory data management-geared advice. Internally, these firms are collaborating with IT, marketing and information security leaders to devise a multipronged big data offering that not only gives clients access to related privacy and security advice, but also exposes them to technology tools and services that they can employ to further their own big data efforts, ethically and within an established risk management context. Instead of jumping on the big data

“In order to survive the new marketplace, law firms must differentiate themselves with something other than the excellent lawyering mantra. Leveraging big data is a way to do that.”

Justin Ergler, Director, Alternative Fee Intelligence & Analytics, GlaxoSmithKline
Big Data in Legal: A Primer

Return-on-investment: All big data efforts should be well scoped and championed internally. The ultimate success is based on the ability to show a hard (measurable) return on investment. This can include expanding your book of business with existing clients (the percentage of profit gain based on big data); producing work more efficiently and at a lower cost (hours/time saved by working faster with more accurate, timely data); or demonstrating defensible disposition and deletion of data (resulting in cheaper and legally defensible storage and data disposable options).

Big Data in Legal: Law firms don’t have a big data problem as much as they are challenged with what a big data world means to them: i.e., their clients’ big data issues. Firms will never reach a scale/capacity issue like government or Fortune 500 companies, but they will be expected to offer sound advice on data security, ethics, overall compliance and risk aversion as it relates to big data. A few U.S. firms have taken the first step and opened separate practice groups focused on offering compliance and regulatory data management-geared advice. Some, such as Am Law 100 firms Fox Rothschild LLP and Kilpatrick Townsend & Stockton LLP have strategically leveraged big data expertise to advise their clients, especially those with increased data exposure risk (such as those in the health care, insurance and financial services industries), on establishing legal and ethical information governance programs in the context of big data privacy and security.

Big data and technology: Big data is not always or primarily associated with technology, but it has potential if legal application and service vendors band together to re-educate the legal community on business intelligence and analytics tools. No lawyer gives a hoot about online analytical processing cubes and data decision points, but providing them insight into their business with meaningful metrics will garner a favorable response. The same goes for relating the corporate big data challenge in terms they feel good about sharing with others. Use technology as an aid, not a crutch, and start with minimal workflow disruption, a relatable interface, and the ability to share and collaborate with colleagues and clients.

Leadership: Gartner predicts that by 2017, the chief marketing officer will control corporate IT budgets and will be a key stakeholder in the big data equation. If this is remotely true, will this potential seismic shift force marketing and IT as well as compliance to all play nice and work together? Can IT help marketing push the envelope when it comes to data mining and client profiling by designing more useable analytics tools? What role will the chief administrative officer/executive director play?

Governance: Developing a governance framework applies to all data, not just the so-called big data sets. This involves setting up rules and regulations as it applies to the data: What can you keep? What are the privacy/security implications? When is the data obsolete or old? When is it ethically inappropriate to keep? Utilize ARMA governance maturity models and frameworks, and focus on governance at the beginning of the project, not as an afterthought.