

OCTOBER 2012

An ALM Publication

# LTN

## LAW TECHNOLOGY NEWS

# DEFENDING BIG DATA

Corporations eager to exploit customer data risk alienating consumers and regulators with privacy and security blunders.

**Mark Melodia**  
Reed Smith

# Defending BIG DATA

Corporations eager to exploit their massive customer data risk alienating consumers and regulators if they stumble on privacy and security.

By Monica Bay

**WE'VE ALL EXPERIENCED THE "ICK" FACTOR** — that queasy feeling that a company has a *bit* too much information about you. Sure, you love that Apple's Genius has figured out what music you like and recommends artists you haven't yet discovered. Yes, you tolerate eBay's suggestions based on past purchases (who can't use another baseball jersey?). Maybe it was just a tad creepy that Netflix offered a slew of crime procedurals shortly after your two-day marathon of *Law & Order: Criminal Intent* while recovering from a flu bug. (See "Why Netflix Thinks I'm Gay," [bit.ly/LTN1210b](http://bit.ly/LTN1210b), and *In re: Netflix Privacy Litigation* [www.videoprivacyclass.com](http://www.videoprivacyclass.com).) It was definitely over-the-top to discover your picture in a Facebook advertisement for a product you "liked" ([bit.ly/LTN1210e](http://bit.ly/LTN1210e)).

But did you know that Target's algorithms can determine with astounding accuracy that you are in your second trimester of pregnancy — because you started buying scent-free lotions, washcloths, hand sanitizers, and cotton balls? And that Target then can tailor the advertising flyer that is sent to your home to include coupons on baby food, diapers,

and other necessities of a newborn? You might find that downright invasive, especially if you are a high school student who hasn't yet been exactly candid with her father. Your subsequent upset stomach may be triggered by something more than morning sickness. Makes you want to actually read those Terms of Service agreements, right? ([bit.ly/LTN2012f](http://bit.ly/LTN2012f))



Mark Melodia  
Reed Smith

Charles Duhigg, an investigative reporter with *The New York Times*, wrote about the Target data project in his new book, *The Power of Habit: Why We Do What We Do in Life and Business* [bit.ly/LTN1210c](http://bit.ly/LTN1210c). Duhigg, who will present the January 31, 2013 keynote address at LegalTech New York, details how Big Data — freely provided by customers — is a gold mine of knowledge about consumer habits that can be used to influence future behavior of both consumers and the companies. Retailers, financial institutions, healthcare providers, insurance and pharmaceutical companies are just some of the corporations that keep tabs on how (and when) we spend our money. Just think about how much data we provide every day, via credit cards, airline elite memberships, health provider records, bank accounts, and all those affinity program tags that hang from your keychain.

---

## FACEBOOK DAILY STATS

- » 2.5 billion content items shared
- » 2.7 billion “likes”
- » 300 million photos uploaded
- » 500+ TB of new data ingested into databases.

— Source: Facebook via Gigaom

---

So what exactly is this new buzzword? “Big Data is an imprecise term increasingly used to characterize the escalating accumulation of data — especially in data sets too large, too raw, or too unstructured for analysis using conventional techniques,” says Paul Bond, a partner at Reed Smith and member of its data security, privacy, and management practice group. Today, humans “create 2.5 quintillion bytes of data each day — an amount so large that 90 percent of all the data in the world has been created in just the past two years, explains IBM on its Big Data website ([bit.ly/LTN1210h](http://bit.ly/LTN1210h)). That’s the equivalent of the content that could be stored on 575 billion 32 GB Apple iPads, says data center ViaWest. (See “The Relative Size of Internet Data” [bit.ly/LTN1210j](http://bit.ly/LTN1210j).)

IBM, active in Big Data long before it had the moniker, offers extensive hardware and software to support data collection, mining, and analysis. “Data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals, to name a few,” the company explains.

IBM defines four dimensions of Big Data:

- Volume (e.g., if you convert 350 billion annual meter readings you can better predict power consumption).



- Velocity (time-sensitive processes, such as catching fraud, need fast analysis of data as it streams to maximize value).

- Variety (structured and unstructured data can include text, sensor data, audio, video, log files, and more).

- Veracity (one-third of business leaders say they currently don’t trust the information they are using to make decisions). “Establishing trust in Big Data presents a huge challenge as the variety and number of sources grow,” says IBM.

Other key technology players in corporate Big Data management include Oracle, Intel, EMC<sup>2</sup>, and SAS, among others.

If you think that Big Data has suddenly gone viral, you aren’t off the mark, say lawyers in its trenches. “In some ways Big Data — and related privacy and security issues — is brand new. But in many ways, it is part of an evolutionary path that the profession and many of us individually have been on for a couple of decades,” says Mark Melodia, co-chair of Reed Smith’s practice group. “Privacy, secrecy, and information security have always been professional obligations for a lawyer, part of our oath and part of our tool kit.”

For legal professionals, Big Data’s tipping point was the May 18 initial public offering of Facebook, asserts Melodia, based in Princeton, N.J. The IPO “drove home to even casual observers the

increasingly close relationship between data collection and corporate value,” he says. At the time of the IPO, Facebook had almost 1 billion users; “the activity of the users is the main asset of the company.” The massive amount of data generated 24/7 on the site’s pages and walls “not only helps Facebook’s advertisers target their ads, this Big Data is a considerable commodity in and of itself,” says Melodia. “Facebook seemed to acknowledge as much by changing its privacy policy to a data use policy ahead of the IPO,” he said. “Shareholders will demand that public companies look to monetize all the personal data they collect, to the full extent the law and public sentiment will allow.”

But the mighty can be vulnerable, he cautions. “Companies that have risen in value on the wings of Big Data can be equally subject to a dramatic fall should data collection, ownership, and use become stymied in red tape and litigation,” he says.

A secondary driver was the “tectonic legal and political shift” to consumer rights that began in the 1990s, Melodia says. That earthquake included new privacy laws, and consequent security obligations, such as the Gramm-Leach-Bliley Act (requiring financial organizations to safeguard sensitive information and explain data sharing), and the Health Insurance Portability and Accountability Act of 1996 (protecting health information), Melodia explains. Then it “exploded,” he says, “with state breach notification statutes (starting in California and spreading in record time to nearly every state). [That] set the stage for the current ‘frenemies’ relationship between much of Big Data and its customers,” he says.

---

## Big Data is an imprecise term that is increasingly being used to characterize the escalating accumulation of information.

---

Another growth factor may be inertia. Storage is cheap these days, and many businesses have realized that it is easier to buy additional servers and not dispose of data, instead of deciding what data to archive and what to destroy, observes Jonathan Redgrave, of the eponymous law firm based in Minneapolis. The firm has nine full-time and two part-time lawyers. “With the simultaneous improvements of algorithms, analytical tools, and processing power, as well as the wide-scale affordability of software, businesses began leveraging vast repositories of data to seek competitive advantages.

“As organizations culled through this data, various privacy and data security questions arose, which led to appropriate security controls and their implementation. Of course, issues relating to legal

discovery and investigations follow in short order,” says Redgrave.

Just as in electronic data discovery, there are also concrete dangers in keeping legacy data, such as enhanced risk that “smoking gun” information might be revealed that would have otherwise been benignly destroyed in the course of established retention policies. (See “Girding for Battle,” page 43, and “What Lurks Within,” *LTN*, Dec. 2011.) This year’s dominant EDD topic, predictive coding — aka technology-assisted review — provides just a hint of Big Data’s capabilities.

Another byproduct of Big Data’s explosion is jobs. (EWeek.com recently reported that there’s a hiring frenzy for data scientists and IT specialists, who can conduct “high-level data analyses and apply it to business projections and modeling.” Top five cities: San Francisco, Washington, D.C., Boston, St. Louis, and Toronto.)

Count in the legal industry, experts say. “There are some really smart partners at top law firms throughout the U.S. who are establishing practice groups for Big Data, information governance, and compliance,” says David Cowen, managing director of New York-based The Cowen Group consultancy. “There will be a need for more lawyers who understand the intersection of Big Data, information governance, and e-discovery,” he said. “And litigation support and e-discovery professionals who get their heads around predictive coding and analytics will be the first to get a phone call.”

**O**N ANY GIVEN DAY, YOU’RE LIKELY TO FIND Reed Smith’s Melodia in court, defending financial institutions in class action suits, the ground zero for trying to find the correct balance between smart business practices that fuel corporate growth and abusing individuals’ privacy and security.

The firm, which ranks 19th on the 2012 Am Law 100, has 1,700 lawyers in 23 offices worldwide. Its Big Data unit “grew from the litigation trenches” and was launched in 2006, says Melodia. Since then, the team has defended more than 70 class actions arising from alleged privacy violations, data thefts and breaches, as well as claims of data misuse involving websites and targeted advertising, he says.

Like other firms, Reed Smith and Orrick, Herrington & Sutcliffe have found that class actions, data breaches, and regulatory activity are the most visible Big Data matters. But the growing consumer pushback has changed class action agendas, observes Melodia.

“For the past two years the focus of the class action litigation has shifted from answering the question, ‘Why did you lose my information?’ to ‘Why do you have my information and why didn’t you tell me you were going to use it to do that?’ ” he says. The wave of class actions “puts directly at issue fundamental questions arising from a Big Data economy:

- “What does a reasonable expectation of privacy mean in a Sally Fields culture in which self worth (and possibly company worth) is determined by everybody desperately wanting to be liked?”

• “Who ‘owns’ consumer information that is shared on websites or in commercial transactions?”

• “Is technology the answer or the problem?”

• “What does ‘harm’ mean in this context?”

• “Can all of these questions be answered by simply writing clear enough disclosures and terms and conditions for consumers to read?” Melodia said.

What has changed in Big Data since the ‘90s “is ubiquity — i.e., the wider range of interaction that average consumers have with the internet, minute-to-minute,” says Antony Kim, co-leader of Orrick’s internet safety, security, and privacy practice group. Consumer involvement includes social networking, powered by strong mobile internet capabilities. That creates “real value” for storing and processing sensitive data (personal or commercial) in networked environments.”

But ultimately, as in most litigation, security and privacy class actions often do not go to trial, Kim notes. “They are either concluded on dispositive motions, or class certification fails, or they settle. And regulator actions, for example, by the Federal Trade Commission, almost invariably end in negotiated consent decrees.”

January 1, 2012, marked the formal launch of Orrick’s “new” practice group, but the team has been in operation since 2009, so you can put it in the veteran column. Orrick, which ranks 16th on the Am Law roster, has 1,100 lawyers worldwide. Kim, based in Washington, D.C., shares leadership with two Silicon Valley partners, Gabriel Ramsey (internet safety and security, litigation), and Stephanie Sharron (counseling and transactions). The unit has 35 lawyers, including three former assistant U.S. attorneys and a former Federal Trade Commission trial lawyer. Team members are based in the U.S., London, Munich, Paris, Beijing, Shanghai, and Tokyo.

They pull lawyers from several practice groups, including intellectual property, litigation, and corporate business. Many sub-specialties are represented, such as technology transactions, emerging companies, employment, insurance recovery. Transactional attorneys advise clients on the technology deals that give life to these new business models; litigators tackle disputes and regulatory investigation, says Kim.

Five non-lawyer professionals — three in IT and network security, an information/data management specialist, and a paralegal — are located at Orrick’s global operations center in Wheeling, W. Va., and can be rapidly deployed to conduct large investigations, forensics and testing operations.

The group’s technology tool kit includes Guidance Software’s EnCase digital forensics suite; AccessData’s Forensics Toolkit; E-fense’s Helix3 Enterprise, a cyber security tool providing incident response, computer forensics and e-discovery tools; ProDiscover Forensics; Palantir, which offers data integration, search and recovery, knowledge management, and collaboration; and IBM i2Analyst’s Notebook, which offers assisted analysis and visualization capacities; among others.

The group’s agenda breaks into two categories:

• Enforcement and proactive measures: Lawyers, investigators, computer forensics professionals, and internet security specialists identify, locate, and prosecute cases against entities involved in internet abuses, including cybersquatting/typosquatting; trade secret misappropriation/corporate espionage; copyright and trademark infringement; online advertising fraud; financial fraud; spam; phishing; malware; and other technical abuses.

• Defensive counseling, advocacy and risk management: With former government attorneys, as well as consumer and HR experts, this group focuses on core internet business issues, including regulatory compliance and investigations; litigation and adversarial proceedings; consumer and employee policies and procedures; data security and breach management; evaluating online revenue models and risk profile analysis; data/records management; and e-discovery.

“Organizations throughout the world — whether they are technology companies whose business models rely on the ability to collect, use, analyze and leverage data, or large multinationals with extensive supply chain and distribution networks — must confront the challenges associated with data privacy, digital security, and internet safety. This is our client base,” explains Kim.

Analytics “can be used to drive traffic to a company’s website, drive e-commerce and advertising revenue, identify trends and patterns of consumer behavior, provide insights into medical and healthcare initiatives, not to mention the diverse array of public policy and educational concerns,” Kim says. “Our clients care about the contractual, legal and regulatory issues that apply to the collection, storage, use, transfer and analysis of these large data sets because data is the relevant currency for our digital world, and reputational and commercial successes hinge on managing data proactively from start to finish.”

Big Data also is “a hot political topic on the Hill,” Kim observes. “Legislators (around the world) are keen to regulate it, enforcement agencies particularly in the U.S. are increasingly getting involved with record-level fines in areas that used to be self-regulated by industry, and the media/blogsphere can’t seem to get enough of the latest Big Data breach.”

Other key practice areas are allocation of rights and responsibilities to data in commercial business transactions, and cybercrime legislation that keeps getting proposed, but not passed, says Kim, who offers startling statistics: The Financial Services Information Sharing and Analysis Center — the trade group that represents the security interests of the financial industry — has reported “staggering” numbers. Since 2005, financial institutions have reported to the Federal Deposit Insurance Corp. and the Financial Crimes Enforcement Network a cumulative \$477 million in consumer loss from online banking fraud, he says. “These are real, reported actual losses.”

“The emphasis in these statutes is public/private partnerships to incentivize threat information-sharing, allowing both companies and the government to better harden defenses and also to better develop enforcement measures against cyber-criminals,” he says. “But balancing features such as liability mitigation for private parties sharing threat data, privacy interests, Fourth Amendment concerns — particularly in the context of the internet, where any forms of control are met with vocal opposition — makes this a long, slow process.”

**CYNTHIA O'DONOGHUE, CO-CHAIR OF REED** Smith's practice group, is based in London. She and Melodia oversee a global team of 40 attorneys — with backgrounds in engineering, software development,



**Cynthia O'Donoghue**  
Reed Smith

cyber security, technology, and in-house experience at banks and insurance providers. The team also helps clients build long-term compliance programs, as well as create contingency plans for emergencies. It also helps clients effectively identify and execute revenue opportunities while respecting obligations to customers.

O'Donoghue leads the international matters. Typical challenges

include privacy litigation and advice, data theft, loss, and unauthorized access. A key difference between the U.S. and international agendas is that “few other countries actually have a mechanism to allow class action suits,” she says. About 90 countries have some type of data protection and/or privacy laws that govern the use of Big Data, she notes.

The group's technology includes smartphones — mainly Black-Berrys, but they test mobile apps with iPhones, iPads, and Android devices. They use Little Snitch Network Monitor for real-time traffic information from applications. For website testing, the team uses a Macbook Pro and runs several virtual machines (both Windows and Apple operating systems) to observe website behavior. They also use Wireshark and Ghostery by Evidon; and Mozilla developer tools to examine information collection and technology deployment.

Paul Bond has noticed a common denominator at his firm: “All of these Reed Smith lawyers share a technology bent in their practices, and, if they had time to look back, nearly all would have to admit that the legal work that they are doing today simply did not exist 10 years ago. The trick, of course, is to figure out what the questions and opportunities will be 10 years from now,” he says. “The future of our economy and society is up for grabs, and now is the time for thought leadership on practical data stewardship.”

**Monica Bay** is editor-in-chief of *Law Technology News*. Email: [mbay@alm.com](mailto:mbay@alm.com). Read more on Big Data: [lawtechnews.com](http://lawtechnews.com).

Reprinted with permission from the October 2012 edition of LAW TECHNOLOGY NEWS. © 2012 ALM Media Properties, LLC.

**ReedSmith**

The business of relationships.™

[www.reedsmith.com](http://www.reedsmith.com)

# International Operations

BY MONICA BAY

Partner Cynthia O'Donoghue joined Reed Smith in January 2008, "with the intention of creating a technology and outsourcing practice in London. "Because a big part of technology relates to the transfer and hosting of data in some form, such as in the cloud, data protection was a big part of my practice," she says. Before she went to law school, she owned a company that offering legal process outsourcing to firms, specializing in writing software to code documents from U.S. discovery for use in litigation.

O'Donoghue's team subsequently merged with Mark Melodia's U.S. operation "to create a global offering, covering all aspects of data protection and privacy, from product conception, to new technology initiatives or rationalisation and the full lifecycle of data — collection, use, dissemination, transfer and destruction, as well as investigations relating to breaches of the national data protection laws by the various E.U. data protection authorities," she explains. "Our client offering includes various toolkits that cover and track data protection laws and amendments worldwide," she says.

Data security is an integral part of data protection compliance in Europe and one of the most important pillars, she notes. Several E.U. countries have "granular regulations on mandatory security standards relative to the type of data being processed, including down to the number of characters that are required for passwords!" Data protection laws cover not just relationships between companies and consumers, but also cover business-to-business relationships, she notes. But the work is not just defensive — Reed Smith counsels corporations on how to generate revenue by using data protection compliance as a competitive advantage, she says.

The practice group also uses project management tools to provide clients well-managed and cost-efficient compliance options. It can be challenging. For instance, one client operates in more than 130 countries worldwide and has a very complex corporate structure — requiring the filing of notifications, registrations, and data transfer agreements with data protection authorities in more than 40 countries and in 17 different languages, she says.

The team also advises clients about "how data can be collected and used as part of new product offerings." For example, for clients in the life sciences sector, the team advises on cloud-computing options that help physicians monitor patients' conditions in real time through the use of sophisticated software-as-a-service — and how the data collected can be "anonymised" and used for secondary purposes, such as research and development for new products or product enhancements, she notes.

O'Donoghue says her experience with software development and U.S. discovery background helps the firm assist clients with

U.S./E.U. cross-border discovery and litigation. "We have developed a protocol that allows organizations to transfer data to the U.S for use in litigation that does not fall foul of the E.U. data protection laws barring transfers of such data," she notes.

## Other current issues include:

- Cloud computing can create regulatory issues relating to the transfer and hosting of personal data — such as data security, issues with transfers of data to locations that may not be known to the cloud customer, and portability and the types of data that can go into the cloud, she says. A hot issue in Europe is potential access to the data by U.S. governmental authorities through the Patriot Act and Foreign Intelligence Surveillance Act.

- Social media sites, the use of mobile apps, and interactive e-commerce websites are targets for investigations by data protection authorities, both in Europe and in Asia — South Korea in particular — especially in relation to geolocation data, says O'Donoghue. Because the use of cookies and other storage devices on websites is now also regulated within the E.U., compliance, data collection, and analytics are issues.

- BYOD (bring your own device) and how employers can keep corporate data secure while providing their employees the flexibility to use their own portable devices is also hot. An area of difficulty relates to the interplay between data security and monitoring the devices that are connected to a company's network — especially if the monitoring software allows the company to actually take over an employee's personal device rather than just keep the company's data sandboxed from the rest of the employee's data and personal use of the device, she notes. "BYOD is an example of how a company can make its Big Data available to employees in a flexible way, without creating additional exposure, and can also involve issues relating to cloud computing especially if employees back up the data held on their devices into the cloud."

- On the horizon are proposed changes to the E.U. Data Protection Directive, which could take effect in 2016, she says. A draft before the European Parliament is broader in scope than the existing directive and would add obligations to appoint data protection officers and conduct mandatory data impact assessments; and create new rights for individuals, such as the "right to be forgotten" and a right to data portability, she notes. The "right to be forgotten" would place a new obligation of accountability onto companies, and shift the burden of proving the legitimacy of keeping and processing data onto corporations by creating a presumption in favor of individuals and their right to control data about them, she explains.

# Big Data Technology

BY SEAN DOHERTY

In the course of reporting our October cover story, “Defending Big Data,” *Law Technology News* sent out a request for information to legal technology vendors that participated in LegalTech New York 2012. We acknowledged the emergence of Big Data and the tension that exists in mining, exploiting, and monetizing customer data versus the security and privacy of that data. We then asked the vendors “if they have launched — or will be launching — products and services addressing Big Data?” *LTN* received a number of responses that show that, for the most part, that legal technologies addressing Big Data are focused on the production of evidence in litigation and government investigations, and not the extraction of customer or consumer preference.

Paul Bond, partner at Reed Smith, said in the cover story that Big Data “is used to characterize the escalating accumulation of data, especially the data sets too large, too raw, or too unstructured for analysis by conventional techniques.” Bond’s reference to “conventional techniques” is to the use of relational databases — which do a great job managing uniform data in data sets that we *thought* were large, until Big Data, he said.

As the standard for big data sets gave way to Big Data, relational databases mostly couldn’t keep pace, said Bond. “The profound coordination that relational databases offered turned from asset to liability. They couldn’t scale fast enough [or] adapt quickly enough to the chaos that comprises Big Data.” As a result, we see businesses and service providers turning to NoSQL and newer data management programs with less need for unified infrastructure, he said. “It’s like we moved from a Newtonian universe of data organization into something more local and relativistic, and the story is just getting out.”

As Bond said, Big Data is moving products and services to satisfy “local and relativistic” customer needs. And legal technology customers have different needs in Big Data than manufacturers and distributors of consumer products and services such as Amazon, Netflix, and Target.

Jeremy Pickens, senior applied research scientist at Catalyst Repository Systems, speaking for himself and not Catalyst, believes that electronic data discovery “is not amenable to many of the approaches that are currently in Big Data vogue.”

“The whole premise behind Big Data is that you have thousands or millions of users engaged with a particular system,” said Pickens. Each user repeating (perhaps with slight variation) the same task, he explained. “From this repetition and variation you can extract information and make predictions about what future users will do.”

Brian Kawasaki, executive vice president of technology solutions at Advanced Discovery, which offers investigatory and litigation services, was not sure if his company “gravitated to truly non-EDD” opportunities

to handle Big Data. Advanced Discovery manages ESI to output evidentiary data, said Kawasaki.

If Advanced Discovery’s e-discovery technologies were re-worked to capitalize on Big Data, the company “wouldn’t necessarily have a good story there,” said Kawasaki. But if Big Data is defined in terms of extracting evidence from volumes of data, said Kawasaki, then Advanced Discovery can build ways to do that and make corporations feel more in control of their data from an e-discovery perspective.

Although Big Data is handled differently in the legal sector than the consumer sector, the technologies used to manage and mine all that data can share the same challenges. According to Robert Miller, of the Rise Advisory Group, there is a growing interest among corporate buyers to understand and manage Big Data. The challenge to date, according to Miller, has been threefold: what technology to use; where to apply it; and whether the organization has sufficient expertise to implement and monitor the software.

Most Big Data products and services view the data from a reactionary position, said Miller. Organizations attempt to make sense of data in the context of an event such as litigation, in investigation. Miller, however, believes that there is an opportunity to move data classification processes to the far left of the Electronic Discovery Reference Model using machine learning to classify and organize data into predefined business use cases near the point of creation. Then organizations can proactively manage information for “retention, security, audit, legal, and business intelligence purposes.”

In *LTN*’s July edition, in “Can Computers Predict Trial Outcomes from Big Data,” Tam Harbert profiled Daniel Katz, an assistant professor at the Michigan State University College of Law, who has been exploring how Big Data can be used by corporate law departments to not just predict outcomes of disputes, but also how to craft strategies, and to decide “whether, how, and where” to file lawsuits. Tymetrix, part of Wolters Kluwer Corporate Legal Services, has accumulated \$25 billion in legal spending data, and Tymetrix has been using analytics to mine that information. One product is already on the market: the \$2,500 *Real Rate Report* that benchmarks law firm rates, and identifies the factors that drive them, wrote Harbert.

Zylab says its e-discovery and information management software has an application programming interface to tap Big Data analytics, but the company is staying “close to home” [legal industry]. We did, however, receive a number of responses that indicate legal technology can be used to meet Big Data customer needs. (Zylab has been heavily involved with international tribunals, such as the European Human Rights Court that just completed a major upgrade of its pub-



lic database that included the adoption of SkyDox. See “The Right to Know,” also in our October issue.)

## E-DISCOVERY MEETS BIG DATA

Denver-based Catalyst, a provider of document repositories and case collaboration systems, manufactures the Insight e-discovery product, said CEO John Tredennick. It is based on the MarkLogic NoSQL Database platform, one of the leading “Big Data” search engines on the market, he continued. Customers routinely handle petabytes of data, he explained, and MarkLogic has tested searches as large as 1.5 million characters and clusters of data can exceed 50 million documents. With MarkLogic, Catalyst’s e-discovery platform supports XML, which enables it to combine metadata, tags, and text in a unified, searchable data store. Catalyst’s technology promises to combine Big Data with a management and analysis platform, Insight, that was not available in a meaningful way prior to Big Data consolidation, Tredennick said.

Shaheen Javadizadeh is vice president of product management at Datacert, a provider of e-billing products to corporate legal departments and their law firms. Datacert has released two products that address Big Data: Legal Data Warehouse and proactive predictive modeling, he said.

Datacert’s Data Warehousing product consolidates case data, time lines, costs, and case outcomes to help law departments better understand the business of law and law office operations and use data to improve results, he explains. The warehousing product also presents law departments with reports and dashboards that aim to help managing partners make better decisions and negotiate better deals with their firms and the business units they serve. Proactive predictive modeling takes historical information and provides the law department staff with years and terabytes of relevant data as cases update and evolve, he continues. By presenting information in real-time, law departments can make decisions that mitigate risk, reduce costs, or improve business results.

Ann Marie Gibbs, national director of consulting at Daegis, characterizes the Big Data problem “as many faceted.” In the context of e-discovery, she continued, “we are faced with ever increasing data volumes, an increased diversity of data sources, and the demand to process data rapidly to meet unrelenting deadlines.” Add to this the need to extract “meaningful” content from data, added Gibbs, which means identifying what you are obligated to produce and what you need to withhold for privilege or other protection.

A Big Data approach to today’s data sets is one that, at its core, tackles problems from a rigorous scientific point of view and relies on statistics to validate results, said Gibbs. This approach is used in Daegis’ Technology Assisted Review, which is currently in Beta (due out this month). Underlying the TAR environment is a scalable Apache Hadoop platform optimized for demanding real-time calculations, according to Gibbs, which is needed to support the demands of machine learning.

Gibbs views Daegis’ TAR product as the first, not the last venture into the realm of Big Data. “We are actively detailing future products and features that will leverage this big data approach to assist our clients,” she said. But, Gibbs reserved, “this technology cannot be applied successfully in the absence of a well thought out process.”

Ted Gary, Exterro’s senior product marketing manager, said that a number of enhancements to Fusion, which is an integrated e-discovery, legal hold, and litigation management software platform, are focused on “greater control and visibility into all of the ‘big data’ that’s created on a daily basis in the enterprise.” Gary added that the upgrades were in direct response to customers who want to ensure compliance with foreign as well as U.S. federal and state laws and regulations, especially those governing data privacy.

Nathan Swenson, director of software-as-a-service development at HotDocs, said the company’s application, called HotDocs Document Services, has some Big Data “tie-ins.” The application lets firms put forms online and sends links to customers. The customers click on the links and they are presented with a HotDocs interview where the users can answer questions and the answers are fed back into the system to publish the story.

The current version of HotDocs does not have many analytical features built into the document creation process. However, said Swenson, the version due out this fall will let the firm or content publisher see metrics about how the client filled the form. A firm can review a report on how far clients got before they abandoned their efforts. Firms will be able to see which dialogs took clients the longest to complete. The goal, assured Swenson, “is to let these people see areas that are difficult or painful for clients so they can have tools to improve their forms and workflow process.”

According to Jennifer Frost Hennagir, director of public relations and investor communications at Huron Consulting, said the consultancy will soon expand its data analytics offering and open its data storage facility in Charlotte, N.C., to address Big Data challenges that corporations are facing. Huron’s data analytics software is designed to further reduce the number of documents in e-discovery and streamline the process that gets at the most relevant documents sooner, rather than later, she said.

At the end of July, StoredIQ announced its newest data intelligence application, DataIQ. Jacqui Galow, director of marketing, the new product is designed to serve as a Big Data ‘start button’ for enterprises. The company pursued the new product after hearing from customers who were not comfortable beginning an e-discovery project — because they did not know the type and extent of their unstructured data without moving it into a repository. (StoredIQ is designed to provide customers an understanding of their data so they can engage in e-discovery and defensible deletion.)