Corporations Turn to Technology to Tackle Legal Challenges of Big Data

By John Tredennick, Founder & CEO, Catalyst Repository Systems

Businesses are drowning in data. The numbers are staggering. U.S. companies create enough data every year to fill the Library of Congress 10,000 times over, estimates say. Walmart alone collects more than 2.5 petabytes of customer data every hour. From a legal vantage point, this amassing of “big data” presents big challenges, particularly when a business becomes involved in a lawsuit or government investigation. In legal matters, rules of discovery require the business to find and produce relevant corporate documents. Big data makes that a Herculean task.

It has helped transform discovery—once a manual process performed exclusively with hard-copy documents—into the most expensive and time-consuming phase of any lawsuit, and that means corporate legal departments increasingly depend on sophisticated technology to sift through mountains of electronic evidence, a process we now call electronic discovery or e-discovery.

Big Data Means Big Discovery

Even before the era of big data, as the complexity of cases and quantity of documents expanded, lawyers were turning to technology to help search and manage it all. Hard-copy documents were scanned into image or PDF files for sharing and ease of use. Later, they were made searchable through OCR.

Eventually, various proprietary platforms, locally installed, came along to manage this process. But as data volumes increased, these legacy systems choked, lacking the power, capacity and speed to handle the quantities of data now common in corporate legal matters. At the same time, corporate legal departments were under increasing pressure to control costs and achieve greater cost predictability—no easy task in the era of big data.

As a result, corporations have been moving to a series of e-discovery platforms and tools that are better suited to the demands of big data, particularly when it comes to costs. The first of these moves was to the cloud. The most obvious advantage of the cloud from a cost standpoint is that it allows offsite hosting of data, eliminating the need for corporations to purchase and maintain hardware, software, and upgrades. Moving to the cloud also requires less IT staff and reduces system downtime, which translates to even more savings.

Apart from cost, cloud systems offer other significant advantages for e-discovery. Foremost among them are the flexibility and scalability to
accommodate legal matters of any size and the sheer power to accurately search, sort and analyze massive quantities of data.

### Hosting Multiple Legal Matters

The move to the cloud helped in bringing a second stage in the evolution of big data litigation technology, the multi-matter repository. This move is more disruptive of the accepted discovery model, which has traditionally been performed on a case-by-case basis. That meant that the process of collecting, reviewing and producing documents was confined to the matter at hand, and that all the work that went into it was shelved once the matter was complete.

But there is an inherent inefficiency in that case-by-case approach. Most corporations are involved in multiple legal matters and typically have large numbers of core documents that may be relevant across cases. Once a document is loaded, processed, and coded in one case, it makes little economic or practical sense to repeat the process with the next case. Thus, as corporations moved their cases to the cloud, the logical next step was to consolidate the cases within a single repository. This benefits the corporation in several ways.

Most directly, it eliminates the duplicate costs that result from redundant processes. In multi-matter discovery, processing, storage, conversion and redactions, as well as some coding, is performed only once to a document or file. The corporation retains the value of its investment in the legal and technical work done in each case.

A less obvious benefit is that legal departments can gain greater control over and understanding of all of their legal matters, leading to more predictable budgets, improved administration and reporting, more efficient workflows, faster ramp-up time on new cases, and the ability to gain legal insight into individual matters earlier, before costs pile up.

### Technology Assisted Review

A final type of legal technology that big data has made essential is technology-assisted review (TAR), which is driving significant changes in e-discovery.

Discovery has traditionally involved eyes-on review of every document. But when a single lawsuit can involve terabytes of data, human review quickly exceeds reasonable limits of time and cost. This explains TAR's rapid rise from convenience to necessity in many cases.

Using TAR, lawyers can algorithmically eliminate the need for human review of large percentages of a collection—often more than half and sometimes as much as 80 to 90 percent. In big data cases, that means substantial savings on review costs—typically, the most expensive phase of e-discovery—potentially shaving millions off a company's annual legal spend.

### Cloud is the logical step to consolidate cases within a single repository

**Next Steps for Litigation Technology**

With the increasing use of TAR have come intensified efforts to further develop and refine the technology. Already, major advances have overcome the shortcomings of first-generation TAR systems. For example, many TAR systems require a senior attorney to “train” the system, adding to costs and often causing delays. Also, early systems required legal teams to have collected all the documents at the outset. This rarely happens in actual litigation, where documents typically arrive in rolling batches.

Recent advances in TAR eliminate these requirements. Through a process known as “continuous active learning,” the team can simply begin reviewing documents as batches are uploaded and the system will continuously learn from their coding calls, continually refreshing relevance rankings as the universe of data grows. The smarter the system gets and the better the results, the more the time and cost of review go down.

Big data is forever changing the legal landscape for corporations. As corporations and their counsel face new challenges from growing data volumes, technology is providing new and better ways to meet those challenges.