

# Legal Perspective

## Controlling Electronic Discovery Costs: Cutting "Big Data" Down to Size

By Brian E. Calla, Scott D. Cessar, and Christopher R. Opalinski

Big data is one of the hot electronic discovery buzzwords of 2012. Big data describes the growing volume, variety, and velocity of information that exceeds the processing capacity of conventional database systems. Some real life examples of big data are:

- 10,000 payment card transactions are made every second around the world
- Walmart handles more than 1 million customer transactions per hour
- 340 million tweets are sent per day. That's nearly 4,000 tweets per second!
- The Radicati Group, a technology market research firm, estimates that by 2013, 507 billion e-mail messages will be sent each day.

Big data poses challenges for litigants by increasing the already expensive process of e-discovery, requiring an even bigger solution. Corporations spend millions of dollars to preserve and analyze huge amounts of data in order to locate information that is responsive to discovery requests and to isolate the relevant electronically stored information (ESI). As the associated costs of managing this data continue to grow, e-discovery vendors are developing new tools and best practices to help corporations manage the ever increasing amount of data involved in discovery.

Earlier this year, the RAND Corporation Institute for Civil Justice (ICJ) completed a study entitled "Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery." The study addressed "one of the most persistent challenges of conducting litigation in the era of digital information: the costs of complying with discovery requests, particularly the costs of review." The ICJ found that the cost of document review is approximately 73 cents of every dollar spent on ESI production. The document collection

and processing phases of complying with discovery requests represent about 8 cents and 19 cents, respectively.

The key to reducing the costs associated with ESI review and production is to reduce the number of documents involved in the process. Two primary ways to reduce the number of documents involved in the process of document review are to develop a defensible document retention and destruction policy and to take advantage of predictive coding technology. These methods will reduce the number of documents that need to be reviewed by attorneys thereby decreasing the overall cost of complying with discovery requests.

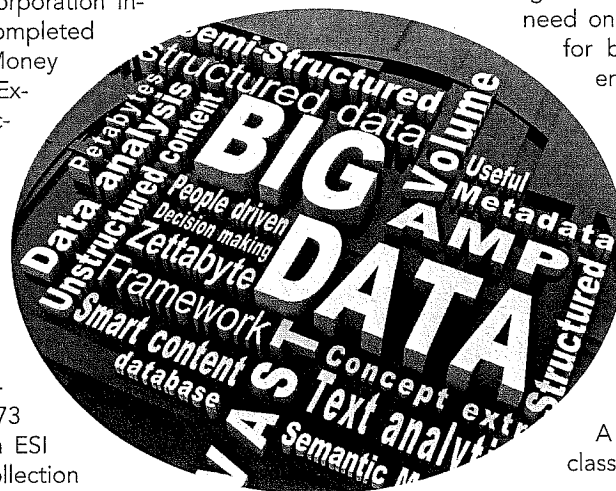
Developing a document retention and destruction policy that is actively enforced and audited is an effective way to reduce the number of documents involved in ESI production. The policy should define the use and storage of not only common storage media such as mainframes, servers, personal computers, backup tapes, etc., but also technologies such as smartphones, instant messaging, and social media.

A defensible document retention policy that addresses retaining and deleting ESI should achieve three goals:

- 1.) Preserve business records while they have a useful life.
- 2.) Provide a defensible explanation as to why certain documents may no longer exist in the event that litigation does arise after documents have been deleted.
- 3.) Limit the number of areas where ESI may be stored (thereby making the process of gathering ESI to comply with discovery requests more efficient and economical.)

Outside of industry specific regulations and litigation hold requirements to preserve information related to ongoing or reasonably anticipated litigation (as well as government investigations or financial audits), a company need only keep ESI as long as necessary for business purposes. For example, emails relating to a construction project should be retained for the duration of the project, but once the project is completed, the documents have served their business purpose and can be deleted. Remember, once information is subject to a litigation hold or if you reasonably anticipate litigation, the information must not be deleted.

A defensible retention policy will classify information in accordance with



a retention schedule that dictates how long each record classification should be kept and when it can be destroyed. The retention schedule should reflect a reasonable document disposal plan that serves legitimate business needs. The policy will also contain citations of applicable document retention regulations in order to ensure compliance with regulations and industry standards. Once a document retention policy is in place, it is critical for the organization to strictly and consistently follow the policy. Strict adherence to the policy is key to the defensibility of the process. The business will always want to be in a position to demonstrate how the elimination of documents is in compliance with a reasonable document destruction plan that serves legitimate business purposes.

Another hot electronic discovery buzzword of 2012 is Technology Assisted Review (TAR). TAR is also referred to as "predictive coding" or "machine learning." The use of TAR is also an effective way to reduce the number of documents involved in an ESI production. TAR is the use of computer technologies to categorize an entire collection of documents as responsive or non-responsive to the litigation based on human review of only a subset of the document collection. As the ICJ categorized it, predictive coding allows computers to do the "heavy lifting" in document review by reducing the number of docu-

to decide whose preferences are "like" yours and which products are "like" the ones you have viewed online.

In the context of the review of a document collection, TAR ranks documents according to the likelihood that they will be responsive to a given request for production. The ranking is based on how the documents have been categorized by the attorney who reviews the documents. Through an iterative process of "learning" from the attorney's categorization of documents as being responsive or not responsive to the litigation, the predictive coding system feeds the likely responsive documents to the reviewer. As the review progresses, and the attorney continues to make responsiveness decisions, the documents that the machine has identified as being non-responsive are set aside (much like when your email system identifies emails that you do not want to read as "spam" and sets them aside.) The non-responsive documents will not be reviewed. Since the percentage of non-responsive documents in a collection can be as high as 70%, predictive coding can provide real cost savings by eliminating the need to review the majority of the documents in the collection. The responsive documents in the collection will be identified through the process of predictive coding, reviewed by the attorney, and produced to opposing counsel. The remaining non-respon-

*Two primary ways to reduce the number of documents involved in the process of document review are to develop a defensible document retention and destruction policy and to take advantage of predictive coding technology.*

ments that must be reviewed by attorneys, thereby reducing the overall cost of document review and production. Although the technology for machine learning has been around for quite some time, it has only recently gained momentum in the context of reviewing documents for ESI production. Earlier this year, we saw the first two cases where the Court approved of the use of predictive coding technology: *Da Silva Moore v. Publicis Groupe*, No. 11 Civ. 1279 (ALC) (AJP), 2012 U.S. Dist. LEXIS 23350 (S.D.N.Y. Apr. 26, 2012) and *Global Aerospace v. Landow Aviation No.*, CL 61040 (Va. Cir. Ct. Apr. 23, 2012).

If machines making responsiveness decisions on documents sounds like a complicated proposition, you are correct. The technology behind it is complicated, but the implementation of it is not. In fact, it is likely that you experience the use of predictive coding technology in everyday life when your email account identifies emails that are likely to be spam and it filters these emails for you. Also, when you shop or browse online, you experience predictive coding. Predictive coding is how a retailer takes what it learns about you with every website visit and uses the information to make predictions about what you might want to buy based on what others with similar preferences have purchased. The program uses predictive coding

sive documents will be sampled to ensure accuracy, but the majority will not be reviewed.

In addition to the potential for tremendous cost savings, there are other possible benefits to using TAR. Studies show that predictive coding is possibly more consistent and accurate than review by humans.

The rapid growth of the amount of electronically stored information that organizations generate has prompted companies to seek ways of meeting their electronic discovery requirements in a cost effective manner. Since document review accounts for the majority of the costs of complying with discovery requests, many of these efforts have focused on reducing document review costs. Developing a defensible document retention and destruction policy and utilizing predictive review technology are two primary ways to help manage the amount of ESI involved in the discovery process, thereby reducing document review and processing costs.

*Brian Calla, Scott Cessar and Chris Opalinski are attorneys in the construction practice of Eckert Seamans Cherin & Mellott LLC. EG*