Information Governance and E-Discovery in the Age of Big Data

Bob Ambrogi

Big data is bringing big challenges to the worlds of information governance (IG) and information technology (IT). Corporations and their legal and IT departments are having to develop new policies and processes governing what data they keep, how long they keep it, and how best to manage it.

The challenge of big data is not merely its volume. Forrester Research is frequently cited for its proposition that big data has four characteristics that push the scale to its extreme – its "four Vs" of volume, velocity, variety, and variability. The concept of big data, Forrester maintains, is not about the data itself, but about the techniques and technologies that make handling data at extreme scale economical.

This is a critical concept but also a variable one, depending on where you are within the IG lifecycle. The Information Governance Reference Model pictured on the next page defines three primary groups of IG stakeholders: business users, IT departments, and legal departments. For each of these stakeholders, the four Vs can have different meanings and implications.

The same holds true when big data moves outside the corporate firewall and into the realm of legal disputes and electronic discovery. In big data e-discovery, just as in big data IG, the challenge is in coming up with techniques and technologies that make it economical to handle data at extreme scales.

In big data e-discovery, just as in big data IG, the challenge is in coming up with techniques and technologies that make it economical to handle data at extreme scales.

The Four Vs and the Power of the Cloud

In the broadest sense of the term, IG encompasses e-discovery. But, whereas IG generally focuses on what happens behind the corporate firewall, e-discovery generally takes place outside it, overseen by outside law firms and using technologies specifically designed for the demands of litigation and government investigations.

When a corporation becomes involved in a legal matter, its lawyers need to dig through the company's electronically stored information (ESI) in search of potentially relevant documents, e-mails, text messages, and the like. This process of collecting, searching, and reviewing ESI is typically the most expensive and time-consuming phase of any legal matter.

In an age of big data, this process would be all but impossible were it not for technology. Sophisticated platforms enable corporations and their counsel to handle the extreme scales of big data with efficiency and economy. These software platforms provide advanced analytics, powerful search, and streamlined workflows to enable legal teams to whittle big data down to scale.

In the early days of e-discovery, before anyone conceived the term big data, these platforms were appliance-based systems. But as big data overwhelmed these legacy systems, corporations moved their e-discovery to the cloud. Cloud-based software platforms are uniquely suited to handling each of big data's four Vs.

1. Handling the Volume of Big Data

Big data cases require big computing horsepower. This is not merely a function of the quantity of data, but also of the increasingly sophisticated search and analytics functions that are being used to help identify and rank potentially responsive documents within all that data.

Cloud-based systems come with the capacity to tap into virtually unlimited processing power. The unbridled power of a cloud-based platform makes it far better suited than a local platform to handling the demands of big data e-discovery. Plus, with greater processing power comes greater speed and efficiency, which translate to lower costs.

In e-discovery, no two legal matters are identical and neither are their data demands. Whereas one case might involve only a few gigabytes of data, the next could involve multiple terabytes. Even within a single case, the data volume can spike unexpectedly.

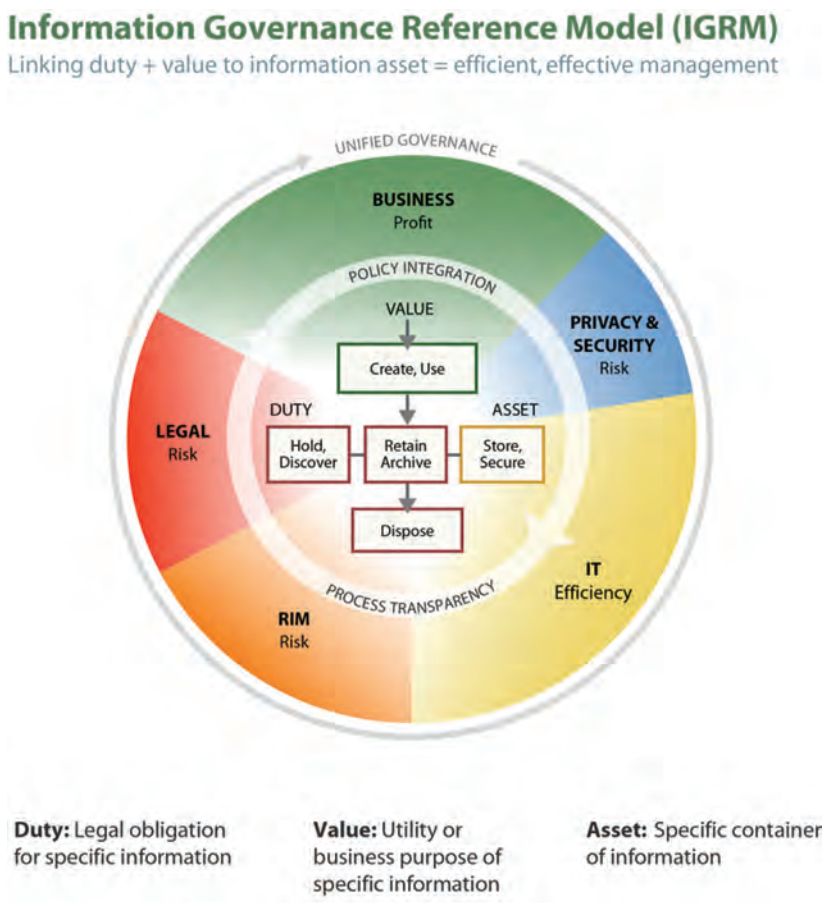
For big data legal matters, therefore, the platform needs to be scalable and elastic in many dimensions. Cloud-based platforms are inherently able to expand and contract as needed and on demand. Users pay for greater bandwidth only when they need it.

2. Managing the Velocity of Big Data

In the context of IG, velocity is not only about the speed at which a business takes in data. In Forrester's analysis, velocity is evaluated in terms of how the accumulation of big data affects the company's decision window for how to process it and what to do with it.

In e-discovery, big data presents a similar decision window challenge. Invariably in e-discovery, lawyers face deadlines to complete the process and serious consequences should they fail to meet them. Their challenge is to review, analyze, and make decisions about huge volumes of ESI within finite time limits.

Cloud-based platforms enhance the ability of lawyers to make these decisions and meet these deadlines in a number of



Information Governance Reference Model/© 2012/v3.0/*edrm.net*

ways. Most directly, they deliver enormous computing power that enables lawyers to use sophisticated search and analytics tools to sift through mountains of data and decide which documents are most likely to be relevant. They also streamline the workflow in which lawyers review and code documents and make final decisions about which to produce to the opposing parties.

3. Handling the Variety of Big Data

In e-discovery, variety has many faces. Legal teams must be able to collect, process, and work with data no matter its original format and no matter where or how it is stored. For global businesses, there are the added wrinkles that the data may be in multiple languages and spread across locations far and wide.

Cloud-based systems make it easy to load and process data from wherever it happens to be – corporate servers, remote custodians, or third-party vendors. Some systems allow users to search data automatically during loading, enabling them to apply predefined parameters to identify and segregate documents up front.

For global legal matters, cloud-based platforms not only facilitate data collection, they also streamline the review process. With a cloud-based platform, review teams can be based anywhere in the world, using many different computing devices and interfaces, allowing lawyers to deploy teams based on language and cultural skills, as well as time and cost considerations.

4. Sorting through the Variability of Big Data

With reference to big data, variability refers to the variable interpretations and analyses that can be drawn from it. In the IG context, this may pertain to data mining and analytics for customer and market intelligence. But in e-discovery, data analysis is critical to understanding the core issues in a case and to properly identifying relevant documents.

To address variability, lawyers are increasingly relying on technology-assisted review, or predictive coding, to identify likely relevant documents and then to rank them by their likely order of importance.

Handling Big Data Economically

In the end, the challenge of big data is in developing techniques and technologies that enable us to handle it economically. That is particularly true in e-discovery. With corporate legal departments under greater pressure to rein in spending, they are increasingly attuned to the efficiencies and savings they can realize through the cloud.

In e-discovery as in information governance, big data's volume, velocity, variety, and variability push the scale to the extreme. But big data does not have to mean big headaches. With greater power, broader scalability, easier loading and processing, and easier access, cloud-based platforms bring to bear advanced technologies that are uniquely well suited to the meeting the challenges of big data's four Vs.

Robert Ambrogi, J.D., is Director of Communications for Catalyst