LAWYERS NEED </ Predictive Coding> TO SUCCEED

Noel Edlin and Erin Poppler

THE BIG IDEA

The question of how to deal with "big data"—data sets of one terabyte or more—is a pressing one for top minds across every industry and academic discipline. From anticipating major shifts in economic patterns to predicting the weather, experts are continually looking for better ways to leverage the glut of digital information we have at our disposal.

Big data refers to analyses you can do at a large scale that you cannot do at a smaller scale in order to extract new insights or create new forms of value in ways that change markets, organizations, the relationship between citizens and governments, and more. The movement toward big data and away from analog information processing has been swift and all encompassing. To provide some perspective, as of the year 2000, only one quarter of the stored information in the world was digital; the remaining 75 percent was analog in the form of paper documents, books, and photographic prints. As of 2007, only about 6 percent of the data were analog; the remaining 94 percent were digital. As of 2013, less than 2 percent were nondigital. If all these data were printed, the paper would cover the entire surface of the United States fifty-two times over. If placed on stacked-up CD-ROMs, they would stretch to the moon in five separate piles!

The legal system and legal services professionals have been relatively slow to adopt new methods for dealing with big data. But our world is changing fast and adaptation is the only way for lawyers—and those who depend on us—to survive and thrive. As big data increasingly dominates the e-discovery process, legal teams need predictive coding, the process by which electronically stored information (ESI) is coded, organized, and prioritized. We attorneys need to start using predictive coding *now* if we are going to succeed at e-discovery and improve the way we prepare for trial.

DOCUMENT PRODUCTION TODAY

The costly practice of overpreservation of documents is one of the key reasons we are facing a discovery predicament. There are too many data, and the costs of storing and processing these data are extremely high. The way document production happens today is not unlike printmaking in the Middle Ages. It is a laborious, linear, manual process. While medieval printmakers were true innovators of their own time, we have come a long way since then, technologically speaking. Yet still we endure the logistical challenges, hosts of paralegals, offsite review, dreaded vendors, high costs, and time delays involved in linear document review. Let's face it—it's not working! Even worse,

From Analog to Digital

The movement toward big data and away from analog information processing has been swift and all encompassing.

Analog and digital data, from 2000 to today:
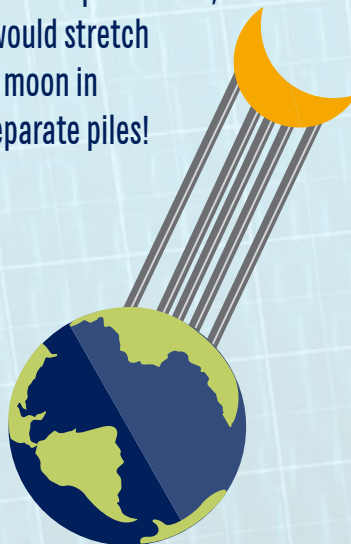
2000

25% digital

75% analog

2007

94%

6%

2013

98%

< 2%

If all the digital data were to be placed on stacked-up CD-ROMs, they would stretch to the moon in five separate piles!

all of these costs are being passed on to the client, a situation that is making it even harder for outside counsel to deliver value to in-house law departments in an increasingly competitive legal services marketplace.

THE WAY THINGS ARE

Let's face it: the legal system is antiquated at best, unwieldy at worst. Like so many other institutions in Western culture it is based on a classical model that was effective two thousand years ago. We would do well to remember that the Greeks were overcome by the Romans and then the Roman Empire—seemingly too big to fail—declined in dramatic fashion. We should also consider that our culture and world have fundamentally outgrown the confines of classical models. Our legal system is inefficient and expensive. We can no longer afford to stick with the status quo.

Change is scary and it requires effort, so most people naturally resist it. But what are the risks of complacency? Consider the situation we currently face in San Francisco with regard to earthquake preparedness. You would think the great 1906 quake and the more recent Loma Prieta quake of 1989 would have provided enough impetus for us to prepare better for the next big seismic event in the Bay Area. But the truth is we are woefully

underprepared and therefore vulnerable to the most terrible outcomes, including extensive property damage and loss of human life.

THE WAY THINGS SHOULD BE

We should all embrace the shift toward using predictive coding for e-discovery. Reliability and efficiency increase when using predictive coding in conjunction with traditional e-discovery methods to review enormous document caches. Law firms that bring the technology in-house can cut client costs by eliminating the need to hire an outside vendor and reducing the amount of time it takes to complete the e-discovery process. Predictive coding is also beneficial because it works well for handling photos, videos, email, and other correspondence—the types of unstructured data becoming more prevalent in e-discovery.

Because the old analog method just doesn't work for many cases anymore, it makes particular sense to use predictive coding in environmental litigation because most cases are data-driven with large document troves; time spans are extremely long; there are various strict liability laws that attorneys must observe; multiple parties are involved; and, due to high costs, there is a premium placed on early settlement.

What Is Big Data?

"Big data" refers to data sets of one terabyte or more. But how much data is that? The following glossary helps put data into perspective:

BYTE
1 Byte = 1 character
100 Bytes = 1 sentence

KILOBYTE (1,000 Bytes)
1 KB = 1 paragraph
100 KB = 1 page

MEGABYTE (1,000 KB)
1.44 MB = 1 floppy disk
600 MB = 1 CD ROM

GIGABYTE (1,000 MB)
1 GB = 1 yard of books
       on a shelf

TERABYTE (1,000 GB)
1 TB = 300 hours of high-quality
       video
10 TB = entire Library of Congress

YOTTABYTE (1 trillion TB)
1 YB = total data stored on
      the Internet

Another reason we would be well-served to embrace the shift toward predictive coding is that courts are increasingly embracing its use. There is an emerging trend in federal courts of accepting—and even encouraging—the use of predictive coding where its use has been at issue in litigation. More and more judges are allowing the use of predictive coding in e-discovery, albeit at a relatively slow pace in comparison to exponentially increasing data sets.

THE KEY PLAYERS

In the past five years, a staggering array of vendors offering predictive coding solutions for e-discovery has emerged. But only a few stand out as real leaders in this new technology space. These include Relativity, Equivio, Axcelerate On-Demand, and Kroll Ontrack. According to a recent survey from the Cowen Group, law firms have had exposure to using more than one predictive coding tool.

A PERFECT WORLD

Data processing will get faster as data sets continue to balloon and the e-discovery process becomes increasingly automated. But this doesn't mean that attorneys will become obsolete. One example comes from the story of what happened in the late-twentieth century when commercial pilots started using more automated flight controls. As automation
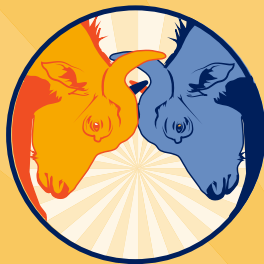
increased, so did plane crashes because pilots essentially forgot to keep flying the plane as they relied more heavily on automatic controls. The lesson here is that attorneys need to keep "flying" the proverbial plane by formulating winning litigation strategies, taking great depositions, and writing compelling motions. This can be done even more effectively as technology increasingly frees attorneys from the time constraints related to linear document review.

In order for predictive coding to truly work, there needs to be a complete transformation in the way we think about the cost-value comparison. When firms take predictive coding in-house we can process data much faster and cheaper. But we also need to hire and train a new class of personnel capable of doing specialized work. Therefore there must be a shift in perspective in terms of what constitutes overhead, as related to traditionally defined roles. Data processing personnel will need to be billable.

We are heading for a more perfect world in which law firms competently handle data processing in-house and bypass vendors. After all, for law firms to continue to exist, there really is no alternative.

Noel Edlin, managing partner at Bassi Edlin Huie & Blum, handles high-stakes environmental, toxic tort, and product liability matters. He served on the California Bar Association's Commission on Judicial Nominees Evaluation and is the former chair of The Bar Association of San Francisco's Judiciary Committee.

Erin K. Poppler, a civil litigation attorney at Bassi Edlin Huie & Blum, has a diverse practice focused on complex litigation, environmental law, and product liability.