IRON MOUNTAIN®

Emerging Trends in Law Firm Information Governance

Unlocking the Power of Big Data, Predictive Coding and 24/7 Access in Law Firms

Introduction

The adoption of effective information governance (IG) has become critical in law firms. Clients remain highly interested in how firms govern their information in light of increased regulatory requirements for personally identifiable information (PII) and personal health information (PHI). Information security has become more complex with lawyers' increased use of mobile devices to access information and the growth of cyber-threats. The Law Firm Information Governance Symposium ("Symposium") was established in 2012 as a think tank to give the legal industry a roadmap for addressing the impact of these and other trends on how law firms operate.

The Symposium Steering Committee, Work Group Participants and Iron Mountain are pleased to provide the legal community with this *2013 Emerging Trends in Law Firm Information Governance* report, a tandem report to the *2013 Building Law Firm Information Governance, Prime Your Key Processes* report. In this report the authors offer insight into three trends impacting law firm information governance: big data, predictive coding and the 24/7 law firm. It is our hope that firms will use these reports to infuse IG into their unique processes and culture, and gain the client service improvement, risk mitigation and cost containment benefits we believe result from effective information governance.

Contents

INTRODUCTION 2
EXECUTIVE SUMMARY 4
BIG DATA AND WHAT IT MEANS TO YOUR FIRM 6
PREDICTIVE CODING FOR INFORMATION GOVERNANCE 16
USING INFORMATION GOVERNANCE AS A FOUNDATION FOR A 24/7 LAW FIRM 23

Executive Summary

Highlights of the three sections of this report are summarized in this Executive Summary.

BIG DATA AND WHAT IT MEANS TO YOUR FIRM

Because law firms are large storehouses of a wide variety of information, they need efficient ways to mine value and mitigate risk from the vast amount of data they collect — and big data can help them do just that. Topics in this report include:

- Examining the opportunities around big data, specifically how it can help firms maximize the quality and value of services provided, improve firm profitability, understand and secure their data and reduce their data storage and maintenance costs
- Identifying the IG processes that will most benefit from big data concepts, including Information Security, Administrative Department Information, Document Preservation and Mandated Destruction, Records and Information Management and Retention Disposition
- Defining the goals of a big data project and involving the right departments within a firm to carry it forward, such as IT, Finance, Business Development and Knowledge Management
- Learning from big data success stories, such as how one firm harnessed case management data into a workflow system that simplifies its pricing model by breaking down cases into actions — and drives a dashboard clients can click through to monitor cases and stabilize their legal spend

PREDICTIVE CODING FOR INFORMATION GOVERNANCE

Predictive coding is a form of technology-assisted review (TAR) that can be used for applying IG within a firm. It leverages automatic classification and coding functionality to bring order to structured and unstructured data. Topics in this report include:

- Understanding predictive coding and how it can be used to reduce the time and resources required for eDiscovery
- Reviewing practical applications of predictive coding in law firms, such as for back–file and day-forward classification
- Illustrating how predictive coding can improve IG on hard-to-administer repositories, such as shared drives, hard drives, SharePoint®, unstructured email and more
- Using predictive coding to automate fundamental records and information functions around email, retention and disposition, legal holds, security and privacy and more
- Selling a predictive coding project internally by developing a proof of concept and showing how the technology will benefit lawyers and the business
- Training the predictive coding system to be an essential component of an overall IG strategy

USING INFORMATION GOVERNANCE AS A FOUNDATION FOR A 24/7 LAW FIRM

With the growing demand for constant access to information, today's 24/7 law firms are looking for ways to equip attorneys with the resources they require to respond to client needs around the clock — from office to home, hotel to airport. Topics in this report include:

- Analyzing the new technology, processes and personnel structures firms are considering to support attorneys' need to work anytime, anywhere
- Identifying potential IG problem areas for the 24/7 firm, including consumer technologies, personal email, mobile devices, physical records, data privacy and environmental concerns
- Developing a blueprint for mitigating these risks through education and establishment of new policies and procedures regarding data being used outside of firm offices
- Building an IT support structure that can analyze and address users' needs, manage business continuity and leverage technologies, such as physical record imaging, where it makes sense

Big Data and What it Means to Your Firm

Compiled by **Bryn Bowen**, CRM, *Principal, Greenheart Consulting Partners LLC*
and **Brian Donato**, *Chief Information Officer, Vorys, Sater Seymour and Pease LLP*.
Additional contributors are listed at the end of this report.

INTRODUCTION AND DEFINITION

Gartner defines big data as high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. This is the ubiquitous "Three Vs" concept of data that now dominates the landscape. In everyday terms big data can be defined as data sets so large and varied that they are difficult to manipulate and mine using traditional approaches and tools. Over time the definition has broadened to include types of transactional and system log data.

In the law firm environment, the biggest of the big data sets typically exist in the email, file share, eDiscovery and social media systems. We will discuss the range of repositories for both structured (e.g., transactional) and semi-structured (e.g., email) data. Generally, information containing the "holy trinity" of metadata – client, matter and employee numbers – can be viewed as structured in the law firm. "Dark Data" residing on years of backup tapes should also be singled out for special mention, as it can be viewed as a potentially toxic data dump.

A. WHY SHOULD LAW FIRMS CARE ABOUT BIG DATA?

Law firms are primarily in the knowledge business. As most Information Governance (IG) professionals would attest, a great deal of that knowledge lies in the vast amount of stored data found in the Document Management System (DMS), email, practice-specific systems and lesser-known repositories, such as file shares. However, as the volume of data grows and the diversity of data increases, gleaning knowledge from this vast store becomes a staggering challenge.

Big data technologies can help solve problems that have vexed IG professionals for years. Imagine being able to characterize and classify vast unstructured data stores that once required human eyes to sort and tag effectively. Tools and strategies now exist to tackle this big data task. Classifying data can lead directly to reductions in data volumes, moving firms closer to the IG nirvana of storing only relevant records and discarding the rest. Classifying data could also lead to greatly reduced risk, allowing firms to more reliably find relevant information in response to a client request or a legal hold. Less data, better classified, means less information management overhead.

Big data technologies hold the promise of automating classification of large masses of unstructured data. Purveyors of Concept-Based Auto Categorization technology to primarily litigation-oriented legal service providers believe that with training, firms will be able to automatically classify emails and documents regardless of language and structure. This technology is at the heart of the current Predictive Coding wave sweeping the litigation support industry. For more details on this topic, review the Predictive Coding report in the 2013 Emerging Trends Report.

For the business analyst or strategist, big data allows firms to address challenges previously thought to be logistically impossible, such as understanding who knows whom in a large firm, detecting inefficiencies in a practice under price pressure or finding trends in the business that require quick response. Additionally, because big data's strength is finding insights and colorations from diverse sets of data, it allows firms to address new and diverse challenges — with the only constraint being the imagination of those asking the questions.

But for the IG professional, using big data in a law firm is all about managing risk and unlocking the value of the information contained in data repositories.

WHAT YOU DON'T KNOW CAN KILL YOU

Oftentimes structured data stores, such as DMSs, don't classify data deeply enough. For example, few law firms can assert with certainty which documents contain regulated sensitive information, such as Personally Identifiable Information (PII), Protected Health Information (PHI) and credit card information. There are several commercially available products that specialize in finding such data, which is the first step in securing it so you don't get hit with a breach problem under privacy regulations.

WHAT YOU DON'T NEED CAN COST YOU MONEY

The concept of redundant, outdated and trivial (ROT) documents and email is one familiar to IG professionals. The same sorts of analytical software that classify documents can help recommend ROT candidates for removal from the system. The more ROT is found, the more firms can reduce storage, backup, server and similar IT infrastructure costs.

WHAT YOU CAN'T LEARN CAN COST THE FIRM CLIENTS

According to the 2013 Q1 Thomson Reuters Peer Monitor® Economic Index (PMI), demand for legal services dropped for the fifth time in seven quarters, demonstrating that law firms continue to be under tremendous client pressure to reduce their fees. While all law firms are aware of the need to provide lower fees to clients without jeopardizing profitability, very few have a handle on how to do so effectively. One place to start is to analyze your firm's raw billing data to truly determine what particular types of matters historically cost the firm. Some firms are automatically applying task codes to time entries in order to allow more uniform analysis of this type.

But the real potential of big data comes when linking information from the Time and Billing system with information from other systems to find otherwise-undetectable correlations. Imagine examining information from docketing, email and expense recovery systems to draw conclusions about how a process might be reengineered to become more efficient. Law firms looking to provide better value while still making a profit will find these types of approaches to mining big data very attractive.

They'd better. After all, corporate clients have access to big data tools and services, and they can use them to see what law firms should be billing for their services. These service providers use millions of time entries collected from a large number of law firms as part of the normal legal billing process and apply data analytics to provide insight into what clients should expect to pay for particular legal services, in context (e.g., jurisdiction). Read more about this in the case study section later.

One of the strengths of big data technology is that you don't have to completely understand the questions before searching for the answers — which suggests that the applications of the technology are limitless. Potential examples of uses for big data technology in law firms include:

- Analyze email reserves to figure out who knows whom.
- Predict which lawyers or groups may be ready to depart, based on system activity.
- Determine why client demand changes. (Example: M&A activity from a client dropped by 20 percent. Analysis of communication patterns suggests the firm upset the client.)

- Understand the likely success rate of a proposed matter (already being developed by vendors in the space for MedMal cases).
- Examine both internal and external data to forecast a (potential) client's likelihood of buying certain services from certain law firms.

B. THE ROLE OF INFORMATION GOVERNANCE IN BIG DATA

IG principles can certainly help a firm manage big data repositories as other unstructured data sources. The *2013 Building Law Firm Information Governance*, *Prime Your Key Processes* report describes how to build IG into 14 key processes. Below, we look at the subset of those processes that is most likely to be impacted by the big data movement. In general, however, the more that IG-based process can improve the structure and definition of data, the more potentially useful that data becomes.

ADMINISTRATIVE DEPARTMENT INFORMATION

Many of the potential applications of big data in law firms revolve around gaining strategic insights from administrative data. IG should influence the development of processes around how this data is captured and analyzed and how the results are memorialized.

DOCUMENT PRESERVATION AND MANDATED DESTRUCTION

It is hard to predict how law firms will use big data technology creatively, but it seems likely that at least some of these uses will generate records that could be subject to a future litigation hold, protective order or destruction order. IG must ensure that document preservation and destruction issues are considered as new big data systems are implemented.

INFORMATION SECURITY

Big data stores could easily contain confidential data, such as information on mergers, or regulated information, such as PII or PHI. As with most new technologies, the capabilities of big data applications are significantly more advanced than the ability to secure the data contained within. IG must be vigilant about the types of information that may find their way into a big data system, and ensure that the proper security principles are applied.

Additionally, some big data "problems" lend themselves best to cloud solutions, meaning large amounts of data could exist outside of a firm's secure perimeter. Whether the firm is hosting their own data in the cloud, or mining data from social media sites, IG professionals should ensure that proper risk-mitigation steps are in place, similar to any application where confidential or regulated information could end up in the cloud.

RECORDS INFORMATION MANAGEMENT

Big data resurrects the question: "What is a record?" The content of many big data systems (for example, analysis and correlation of events from a variety of system log files) will involve data not found in typical documents. IG professionals will need to wrestle with what definitions of "a record" make sense in the context of big data, and should be aware that the answers will likely vary based on the technology, the application and perhaps even the location of the system.

RETENTION/DISPOSITION

Currently, most big data solutions are more focused on gaining insights from the data rather than on what happens to the data after the process is complete. In this environment, correct disposition time frames for big data repositories are hard to identify and even harder to implement. IG professionals should help push the agenda by ensuring that there are procedures in place to identify the source data for big data solutions, and providing feedback on the controls necessary to properly enact the firm's information lifecycle management processes. Note that this issue could be more complicated if the big data solution is hosted with a cloud provider.

C. HOW TO ENGAGE BIG DATA

As a starting point, define the goal as understanding where structured and semi-structured data exists in the enterprise. Then, introduce the M3 concept of managing (e.g., applying controls like legal holds or ethical walls or disposition guides), mining (e.g., gathering information to answer a question) and manipulating (e.g., presenting results of queries in context) data in place as the logical outcome of successful application of IG principles.

Creating a data map is the first step in identifying where information lives in the firm and understanding its DNA. This can be accomplished by separately indexing content and metadata of data objects in identified repositories, and then displaying them in "heat map" representations. There are a number of commercially available indexing applications that can handle this task. A three-dimensional look at the data can also be accomplished by performing this action over time, with the goal of understanding what it is, who is generating it and under what conditions.

The benefit of applying IG principles to this exercise is to make data reliably available for other interested parties to access. By understanding what data exists and where it is, it becomes easier to understand the context within which it was created in the first instance. As such, it becomes a more valuable resource that is much less susceptible to "noise" and the risk of misinterpretation. In effect, it goes from big data to "data under management." These indexing tools can also be employed to identify data that may no longer be useful to the firm and data that can be classified as ROT — and to apply classification and context to the remaining information.

The role of IG is to make data available for other parties to access and mine in readable, usable form, and to make it available for others to utilize. Records Management is being asked to manage the changing view of what useful information is. It is also being asked to ease its movement from a relational, structured, database-driven world toward a semi-structured, social media-driven world, following the trend across industries served by law firms.

D. WHO WILL JOIN THE BIG DATA REVOLUTION?

A big data offensive should have no trouble finding allies, since a phalanx of different departments within a law firm stand to benefit. For example:

INFORMATION TECHNOLOGY

Most obviously, IT can benefit from a reduction in infrastructure costs if big data is successfully used to remove ROT from the environment. Additionally, IT has problems that analysis of big data can solve, such as correlating the vast amounts of machine-generated logs to detect and alert about significant events, or drawing patterns out of help desk systems to determine areas for new education offerings.

FINANCE

There is a plethora of opportunity for the firm's finance department. As a precursor to big data, as we define it here, Business Intelligence (BI) can be aimed to provide visibility into what drives the financial performance of the business by slicing and dicing financial information. Big data can bring a whole new level of insight. For example, using big data tools to examine the firm's vast store of time-card data could provide a key competitive advantage when pricing services. There are also opportunities to examine what competing law firms are doing with price, hourly rates and even work product using publically available information and information available for purchase.

BUSINESS DEVELOPMENT

If your business development department runs the firm's web site, they will be very interested in what big data can tell them about how the web site is used, how people click their way through it and what patterns most often lead to an inquiry about services. Using internal data about the type of work the firm does for each client, and external data about the variety of legal services other law firms perform for the same client, business development might be better able to guide cross-selling efforts and target resources appropriately. A simpler effort might be to mine the email system to determine which attorneys actually correspond regularly with which client contacts.

KNOWLEDGE MANAGEMENT

There are exciting opportunities for the firm's Knowledge Management (KM) employees at the intersection of Predictive Analytics, big data and Enterprise Search. For example, imagine using the vast information from a firm's DMS activity log to determine which documents are most frequently utilized by attorneys as a starting point for new work product. It might even be possible to analyze the documents themselves and determine which ones are most often the sources of cut-and-paste operations, or to correlate this information with raw time-entry information to determine which document-creation paths produced the most efficient outcomes.

Additionally, KM could compare results available within a variety of docket systems with internal information from the DMS to draw correlations around what contract clauses were the most successfully litigated, or the rate of success of summary judgments based on industry, case type or other classifications of interest. These correlations can help guide future successes.

E. HOW TO IMPLEMENT BIG DATA SOLUTIONS

So far we've been examining the "what, why and who" of big data, but you may be asking yourself, "How do I do all this?" Let's explore some of the available tools and techniques.

BIG DATA BEFORE THE BIG DATA CRAZE

Prior to the current big data revolution, there were many tried and true solutions for pulling large data sets from disparate databases. These tools are now well defined and mature, and they work well to solve some of the problems in the big data solution space.

When the data is structured and there is knowledge about the sorts of queries the business wants to run against the data, standard BI tools, such as Data Warehouses and online analytical processing (OLAP) cubes, are excellent solutions. In standard BI applications, it is necessary to first identify the sources of structured data that are of interest (often from SQL databases), then determine the best way to relate the data. Typically, IT will create an Extract, Transform and Load (ETL) process to automate the population of a data warehouse. As the name implies, the process extracts the data from the source databases, transforms it so it can be queried easily and loads it into a data warehouse platform. Once in the platform, the data is available for query and reporting purposes. Usually, the ETL is run on a regular schedule, but the data is not updated in real time. This process greatly reduces the time it takes to run reports and the impact such queries have on the operation of the source database systems.

In a law firm, client, matter and partner profitability are examples of typical BI applications, and there are several products that can be used to automate the BI process for financial reporting within law firms.

If you have used a spreadsheet, you are familiar with rows and columns. An OLAP cube[1] can be thought of as a multidimensional spreadsheet. A data analyst might set up an OLAP cube on a data warehouse using a tool, such as Microsoft SQL Server® Analysis Services (SSAS), with predefined dimensions that allow the business to drill down on information it needs to know. Using our example above, if the dimensions included partner, location, practice group and industry, the OLAP cube would facilitate a report on profitability by the client's industry, broken by practice group and then by partner. It would also allow a data analyst to drill down into the cube to examine a specific combination of values (e.g., profitability of clients in the ice cream industry with relationship partners in the Alaska office).

PANNING FOR BIG DATA GOLD

Admittedly, BI solutions have their roots in the financial reporting world, and while some of the applications discussed in this paper are financial problems, many are not. Traditional BI solutions may not be the best analytical tools to use if the subject data isn't highly structured and clean, and if there isn't a clear understanding of the queries intended to run against the data. For such problems, a new set of tools has emerged.

[1] Note that "cube" is somewhat of a misnomer because OLAP cubes are often more than 3 dimensions and rarely have equal sides.

THE BIG WHOOP AROUND HADOOP

It is hard to talk about big data without eventually talking about Hadoop®. Hadoop is an open-source framework originally developed at Yahoo to improve the scalability of a search-engine project. The strength of the Hadoop framework is the ability to carve up work into small pieces, assign each piece of work amongst many machines (potentially thousands) and then manage the process of reassembling the pieces as the work completes. The framework is a collection of many technologies, all open source, the most famous of which is MapReduce.

Originally developed by Google, MapReduce performs two sets of functions: The Mapper part of MapReduce is responsible for breaking the data down into something meaningful, which can be expressed as key value pairs. Mapper works on many similarly sized chunks of the data at the same time, which allows it to handle large amounts of data and complex jobs. The Reduce function takes the results of all of the Mapper functions and collates the results.

Hadoop is a powerful and viable tool, especially when combined with cloud computing services, such as Amazon® Elastic MapReduce. In Amazon's cloud, it is possible to create hundreds of instances of a Hadoop node to work on the same job and pay only for the computing time and storage you use.

To better visualize what Hadoop does, imagine a text file that contains 10 million narratives from time card entries. The goal is to determine how many time cards contain a set of keywords, such as "conference," "drafting" and "deposition," and the amount of time charged on each time card. The Hadoop framework divides up the files into chunks in order to take advantage of all nodes available to it. MapReduce then goes to work. The Mapper function runs on each node, crunching the data assigned to that node to determine the count of keywords. The result from a single node, a set of key value pairs, might look like this: (Conference, 2303), (Drafting, 512), (Deposition, 343).

Once all nodes are finished, the Reduce function of MapReduce takes the results from each node and combines them to produce one result set for the entire text file. If the results looked like this: (Conference, 127304), (Drafting, 22512), (Deposition, 155343), you might conclude that the lawyers in your firm should spend more time drafting and less in meetings!

The challenge with Hadoop is that someone with programming skills must write a program to define how to crunch the data and arrive at the key-value sets. These programming functions may be written in Java®, Python® or special big data languages like Pig. For many law firms, finding and training such resources could prove challenging. However, there are alternatives.

BIG DATA ALCHEMY

Certainly, there are a number of companies and consultants that can provide Hadoop expertise, but it is possible to implement big data solutions without using Hadoop or other open-source solutions. For example, several companies offer out-of-the-box solutions aimed at solving specific IG problems.

These solutions break down the big data challenge into basic steps, starting with finding unstructured data and ending with the disposition of all data identified. They frequently offer connectors to common data repositories, such as SharePoint® and Microsoft Exchange Server®, and feature their own proprietary indexing and search engines. By reducing the solution space of big data, and focusing on a specific problem set, these tools offer the opportunity to implement meaningful business solutions in a more manageable time frame — and most likely at a greater cost. Interestingly, many of these solutions can trace their roots back to the eDiscovery provider industry and focus on information processes, such as litigation holds, auto-classification of email and semi-structured information with an IG perspective.

The Splunk® solution focuses on a different aspect of big data. Its platform is aimed at collecting, indexing and making sense of the vast amount of data that computers generate during their operations. Common applications include monitoring for system failures in the environment, examining logs for evidence of suspicious behavior and even an eDiscovery app aimed at machine-generated data. Part of the approach includes an app store full of user contributions that can customize your Splunk installation. There is even an app called "Finding Apps is Hard as He**" which analyzes your data and suggests useful apps based on what other installations are doing with similar log data.

There are many other vendors and tools, including a slew of NoSQL applications. This section was intended to provide a sense of the options available to tackle big data analytics.

A. Case Studies Using Big Data for the Legal Industry

Not surprisingly, the fruit from harvesting and processing big data has already been canned in the legal industry, both by law firms and providers in the space. A few interesting cases are presented in the following subsections.

Littler CaseSmart®

Leveraging the significant amount of data gained from the hundreds of administrative actions (primarily EEOC) handled by the firm, Littler Mendelson created a workflow system incorporating templates that bring efficiencies to the process by breaking down every step in the handling of these actions. The process was captured in proprietary software, which then allows Littler to offer its clients with large numbers of these types of cases an "all you can eat" pricing model.

The result is a big success because it provides clients with the opportunity to stabilize their legal spend. What's more, through a client-facing dashboard that shows the status of each case and allows click-through to every detail, the client is able to analyze cases to spot trends in violations — which can afford them the opportunity to take corrective action before violations occur.

Wilson Sonsini's Convertible Note Term Sheet Generator

Wilson Sonsini Goodrich & Rosati (WSGR) has been the premier venture financing counsel for tech startups and similar companies for quite some time. This expertise and the accompanying trove of information associated with hundreds of these deals positioned the firm to mine and leverage data for future deals. The effort originated as an internal tool for WSGR attorneys to rapidly generate draft term sheets that they would then polish and deliver to clients.

Using the information from an online questionnaire completed by the company seeking financing (usually a tech startup), the Generator creates a venture financing term sheet based on those responses, which can provide immediate transparency to foundling companies in the fast-moving tech sector. The tool also provides an informational component, with tutorials and definitions of financing terms. It is one part of a suite of document-automation tools used by WSGR to generate start-up and venture financing-related documents.

Quantitative Legal Prediction by Lex Machina

The question: "What are the odds of winning this patent litigation?" The answer: Use information from court documents to predict the outcome. Lex Machina, a company with its origins in Stanford University's IP Litigation Clearinghouse, has spent the last 10 years building an effective case-prediction database in patent litigation, which is both a high-value and high-cost area for corporations.

Its database contains information from over 128,000 IP cases and more than 134,000 attorney records, in addition to information on the judges and law firms involved in those cases. It took data scientists, attorneys and engineers over 100,000 hours to conform and normalize this information to make it available for analysis. This effort was definitely worthwhile, since corporations spend significant amounts of money to both procure and protect intellectual property because it is often a significant portion of their value. In fact, according to a report by the Federal Judicial Center, the average cost of taking a patent case to trial is around $5 million per patent. In addition to commercial clients, Lex Machina also makes the information available for use by policy makers, media, academics and other purveyors of the public interest.

Aggregating Law Firm Billing Information for Optimal Pricing for Legal Services by TyMetrix
Starting out as a vendor of e-billing and matter management systems for corporate law departments, TyMetrix created
its collection of data on billings and other metrics associated with legal matters in 2009. With its customers' permission,
it has since accumulated a warehouse of data from over $25 billion worth of legal spending. TyMetrix then uses analytics to mine the information for use in its products.

The TyMetrix Rate Driver mobile app can calculate average hourly legal rates for lawyers across the U.S. based on such factors as law firm size, geographic location, attorney level and experience and area of specialty. The data for the app comes from another TyMetrix product, the *Real Rate Report*, which benchmarks law firm rates and identifies the factors that drive those rates. TyMetrix also offers a free app for mobile devices that uses *Real Rate Report* data to serve up average hourly legal rates of law firms across the country. The ultimate goal of this data is for TyMetrix to provide its clients with the tools to effectively run "what if" scenarios to forecast future cases in order to manage legal costs.

CONCLUSION

At this point, it should be clear that utilizing big data effectively will accelerate the transformation that is already underway in the legal industry. The technology has the ability to impact areas as diverse as correctly pricing value-based matters to finding and correctly classifying dark data and ROT. It may be the only way law firms will solve the data explosion caused by social media, and it could provide a most useful tool in the fight against hackers. And, many firms will be able to realize the oft-promised institutionalization of certain clients through both effective cross-selling and competitive, holistic pricing models.

For big data to reach its potential in law firms, IG professionals must play a key role not only by utilizing the technology, but by ensuring that others observe sound IG practices as they develop big data solutions. With IG leading the way, only the imagination limits where big data can take the legal industry.

APPENDIX A

Checklist for Assessing Data at Your Firm

A. PERFORM A GENERAL ASSESSMENT:

- Inventory data in identified repositories. Use available resources to assess data (i.e., create a high-level data map and categorize data as either structured [identifiable by client, matter, attorney numbers] contained within a database structure or semi-structured). Identify sensitive information (PII, credit card numbers, etc.) and perform corrective actions (delete, mask, protect) as an immediate measure.
- Identify duplicates and perform corrective actions (delete, archive, etc.). (This can be done over time, but should be done before any serious analysis of the data is done.)
- Decide which information should be moved and which can be managed in place in an ongoing fashion.
- Establish a solid categorization structure (taxonomy) to mine and use data more effectively.

B. ADDRESS THE ISSUES RELATED TO STRUCTURED DATA, INCLUDING:

- Examine data integrity. Use heat maps to detect trends in how your lawyers are working and to understand the context.
- Increase your competitive intelligence through data mining. Have records managers assign retention schedules, clean up data and organize data-mining programs.

C. ADDRESS THE ISSUES RELATED TO SEMI-STRUCTURED DATA, INCLUDING:

- Find out where it is, what it is, what shape it is in (quality of metadata and content) and who created it using what processes. Several commercially available eDiscovery focused tools exist which can be deployed to this purpose.
- Determine the level of ROT.
- Have IT work with the RM group to identify what is in the data to prevent it from clogging up the system. (In most enterprises, including law firms, over 40 percent of information is ROT, and of these characteristics, trivial is the most difficult to determine.)
- Find a major pain point and focus on practical solutions, such as disk-space reduction, to encourage partner support.
- Establish change-management procedures by selecting the type of change you want to promote in order to produce more usable information from current processes. Then, decide how to communicate your procedures, set priorities and identify champions (i.e., people who will benefit from both the increased efficiency of the process and the usable information output).

D. FOLLOW THIS GENERAL GUIDANCE AS YOU ANALYZE YOUR DATA:

- Create and apply a solid categorization/structure scheme to use data more effectively.
- Work in new ways by cross-pollinating with people you haven't worked with in the past, in order to understand how and why they create and use information in their particular areas and how it is stored.
- Look at connecting silos of information. Social media may provide connections and better access to information that someone else has created (e.g., use Twitter as a newsfeed, employing hash tags to follow areas of interest). Begin building an ecosystem of shared information.
- Obtain ideas for topics to analyze (or, unasked questions to be answered) from primarily business-focused personnel, such as rainmaker partners, practice group leaders, marketing and business development, etc.

Predictive Coding for Information Governance

Compiled by **Rudy Moliere**, *Director of Records and Information, Morgan, Lewis & Bockius LLP*,
Leigh Isaacs, CIP, *Director of Records and Information Governance, Orrick, Herrington & Sutcliffe LLP*,
and **Samantha Lofton**, *Director of Records Information, Risk Management and Practice Support, Ice Miller LLP*.
Additional contributors are listed at the end of this report.

WHAT IS PREDICTIVE CODING AND WHY SHOULD IT BE CONSIDERED AS A SOLUTION FOR INFORMATION GOVERNANCE INITIATIVES?

Predictive coding is a form of technology-assisted review (TAR) or computer-assisted review (CAR) that is used to train technology on the automatic classification and coding of unstructured data. The technology can be used for applying Information Governance (IG) within a firm to both structured and unstructured data. One key component of predictive coding that differs from searching analytics is the methodology for training the technology that is used to automatically classify records and improve the accuracy and self-learning of predictive coding technology.

Understanding predictive coding and related terms will help with defining its use. In the eDiscovery landscape, predictive coding is referred to as automated review, TAR or CAR vs. Linear Review. Linear Review is human-directed review. In linear-review situations, eDiscovery teams manually analyze documents for key terms and relevance until all have been reviewed within a litigation support review platform. This approach is often used when each document in a collection needs to be reviewed. However, with the emergence of big data and the explosion of electronic documents and manage-in-place repositories, the time, resources and costs required for a manual or linear review have become unreasonable due to the size of data collections, especially when compared to TAR.

For example, what if you were searching a large collection of documents, looking for a single maintenance record that confirmed reasonable due diligence vs. negligence in a failed mechanical system? In this case, TAR could enable a reviewer to code or identify a random sample of documents as relevant, non-relevant or privileged and define relationships to weed out the insignificant documents. The predictive coding system would analyze the remaining documents and determine which documents are responsive or fit in the defined categories based on past experience.

One critical step in a successful predictive-coding process is validating the prediction results. Reports can be generated to measure the accuracy of the trained system. A percentage of the documents reviewed by the computer that are included or excluded are reviewed by the subject matter expert. Any needed corrections are made and the technology is either retrained or it learns the adjustments. The dataset is then reanalyzed and the process is repeated until the desired level of accuracy is reached.

It is important to understand the difference between the analytics of predictive coding and those of predictive filing. Predictive filing works based on established rules, such as "email from x.com goes to a specific client matter." Currently, you can run searches in email and move data to a defined location based on a key word. However, predictive coding takes things one step further by offering the ability to teach technology the context of the information. The technology is then

directed to find relevant information regardless of location. As an IG solution, predictive coding should be considered when looking at email/PST files, shared drives, SharePoint® and firm document repositories. We will discuss later other records and information management (RIM) benefits to using the tool. Predictive coding provides a proactive methodology that can be implemented as part of an IG program to aid in the organization and classification of data for the purposes of protecting the organization.

PRACTICAL AND BENEFICIAL USES AND APPLICATIONS

Predictive coding and analytic technologies are not new. Their usage in eDiscovery has become prevalent, and they are an efficient and cost-effective way to cull an inordinate amount of data. Many organizations are beginning to explore and reap the benefits of leveraging these technologies for IG purposes. Law firms share the same challenges and should closely evaluate the potential of these tools for implementing an IG program.

Traditional technology offerings, while continuing to improve, still often fall short in providing an easy, seamless method of supporting an IG program. For attorneys and staff, the priority is ease of use and accessibility as opposed to information management. Thus, creativity is required to achieve a balance of efficiency and compliance and ensure that all data (including emails) is effectively governed.

Predictive coding and analytics can be used for both back-file and day-forward tasks, and it can assist with a historical classification of information in email and other files. It offers the ability to methodically structure currently unstructured data and apply information policies (back file). Predictive filing utilizes the available tools to support proactive filing of information, or auto-classification, during its lifecycle (day forward). Deduplication capabilities and the ability to identify information as a record within a records management system are key benefits. More advanced solutions include the ability to track what has been deduped.

How can these tools specifically help with IG and traditional records management functions? There are many relevant areas to consider, including IG, RIM functions and overall benefits and considerations.

INFORMATION GOVERNANCE

Unstructured data presents significant IG challenges, and thus far has been a conundrum. Attorneys appreciate the ease of use of these disparate, flexible repositories, even though information managers find them challenging to administer. Predictive coding and analytics tools can be a valuable solution for:

- **Shared drives:** Often associated with the "wild west," shared drives can benefit from predictive coding by indexing and creating a decision tree based upon key words, leading to advanced concepts of information relationships. Subsequently, client/matter tags can be associated with the content to apply the required information lifecycle management processes.
- **Hard drives:** The use of synchronization tools can assist in applying predictive coding on local computers and allow management of key data.
- **SharePoint:** The ease of setting up SharePoint sites (functioning like shared drives) often results in unmanaged, mushroom-like crops throughout the firm. Predictive coding can be used to manage this information in place, or identify and tag it for transfer to a data management system (DMS) or other repository.
- **Unstructured email (a.k.a. the dreaded .pst):** As previously mentioned, data can be categorized as "upstream" and "downstream," which defines how the deployment will be performed. Whether it is capturing historical information or active email, this approach can be used to predictively file email into the appropriate content repository.
- **Any indexible repository:** Due to the challenges of managing them, firm extranets and intranets are sometimes omitted from a governance program. Predictive coding can simplify those efforts so that relevant client and administrative information can be identified, classified and managed here as well.

RECORDS AND INFORMATION MANAGEMENT FUNCTIONS

Predictive coding and tools that aid in auto-classification provide opportunities to automate fundamental RIM functions, and offer innovative methods to apply traditional lifecycle management to the firm's information.

- **Email:** Email has replaced the traditional methods of correspondence, and often contains information that should be captured as an official record. The conversational nature and sheer volume makes email difficult to manage. As discussed above, email can be brought into a governance structure using predictive-coding tools.
- **Retention and disposition:** Due to multiple repositories and unclassified data, achieving consistent, unified and defensible disposition is challenging. Predictive-coding tools offer a potential solution to apply policies during file transfers (incoming and outgoing) and final disposition (such as destruction/deletion or return to client). Common law firm retention and disposition activities are:
  - **Lateral transfer of information (inbound and outbound)**
    - **Onboarding — client file intake:** Importing client data associated with a lateral attorney into the firm's systems should be managed carefully. Established processes and procedures, combined with technology, can offer auditing capabilities and help maintain the integrity of the incoming information. This basic premise also applies in cases of mergers or acquisitions — both during the due-diligence process and post-merger. Predictive coding and analytics can add efficiencies by identifying and coding large amounts of information by client/matter for import into the firm's systems. As a result, they can give incoming attorneys prompt access to critical — and potentially revenue-producing — information.
    - **Off—boarding — client file transfers and attorney information:** Departing partners routinely take clients with them, and it is common for clients to request the collection, production and transfer of their information. This information is often voluminous and composed of a mix of structured vs. unstructured data. When executing these processes, it is often necessary or helpful to separate internal vs. external data. Attorneys are frequently called upon to conduct a review and provide direction on final disposition of content, and these technologies can help organize or reduce the volume of information that must be reviewed.

      In addition, when attorneys leave a firm, it is not uncommon for them to leave behind a trail of personal or administrative files and data. It is often unnecessary to retain this legacy material — and potentially costly to ignore it. Departing attorney data can be identified and analyzed so that it can be captured and retained as "know how," provided to that attorney or destroyed. It can also be used to update core systems of a change in responsible attorney should a client associated with the departed attorney choose to stay at the firm.
  - **Identification of vital and/or historical records:** As with any business, law firms must be diligent in their retention of vital records. Often, legacy information provides a history of the firm and may be important to retain. By establishing specific criteria, these technologies can aid in locating, segregating and/or securing this information.
  - **Disposition:** Once the retention period for client and administrative files has been met, a variety of disposition activities may occur. It is important to execute a unified approach. Predictive coding and analytics can be a valuable tool for accomplishing these tasks in a defensible manner.
- **Legal hold/preservation solution:** The ability to effectively implement and maintain a legal hold is a critical component of any disposition program. By addressing unstructured data, these tools can make data collection easier and help establish criteria for preservation. They can also remediate duplication of data across multiple systems and locations.
- **Security and privacy:** Firms are custodians, not owners, of client information, and have a duty to ensure security and privacy. Predictive coding and analytics can be used to analyze employee behavior to determine suspect activities proactively, such as exporting or emailing client documents outside the firm without prior client instruction. They can also be used to identify and apply policy to Personally Identifiable Information (PII), Personal Health Information (PHI) or other similarly sensitive information for both clients and employees.

- **Other potential uses:** Predictive coding and analytics also allow firms to leverage information in non-traditional, innovative ways.
  - **Time and billing systems:** Using these tools with time and billing systems could enable the use of routinely captured information to train other technologies.
  - **Conflicts analysis and risk reduction:** When engaging new clients and matters, it is critical to identify any potential conflicts. Predictive coding and analytics tools can be used to mine data and search for revealing content that may be critical for the analysis, but not otherwise found. It's equally important to identify information that should not be imported into firm systems. These tools can be used to segregate information for potential clients that have not yet cleared conflicts or been accepted as a client of the firm.
  - **Business intelligence:** By taking an analytical, calculated look at information in key data repositories, relevant information can be extrapolated for use by firm leadership. Leaders can translate that information into strategic decisions, business development opportunities or competitive advantage.

These are but a few use cases. There are no doubt additional opportunities to be discovered from the use of emerging predictive coding and analytics technologies. Regardless of how they are used, however, several key elements should be considered in order to achieve optimal value and return on investment:

- Users must have a clear understanding of business, IT, risk and compliance requirements.
- Predictive coding requires significant attention to set up based upon the initial goals of the search.
- Key objectives for predictive coding could be to assign client and matter number, custodian, document type, etc. to the analyzed documents.
- Predictive-coding tools can be categorized by workflow requirements for their applications.
- With litigation support, the tool must be retrained on a per-case basis. IG setup may be less intensive.
- Additional predictive-coding tools may not be required for email filing and searching. These features are often available as an out-of-the-box capability of email management (predictive filing). It is important to first evaluate what tools are available as an alternative before making an additional investment.
- It is easy to set a "BHAG" (Big Hairy Audacious Goal), however, it is important to remember that good is often good enough. Getting a client/matter number associated with email and other data for retention purposes may be all you need. Carefully weigh the ideal vs. the practical approach. Pulling information from multiple data sets to be loaded into a review tool may require specific and expensive personnel resources to extract and import data between multiple systems. It is important to evaluate the available software and resources to address the internal needs of the firm's IG program.
- It is essential to ensure that policies are appropriately and consistently applied.
- In addition to investigating internal technology tools that can be tapped, it is equally important to remember that there is expertise that can be leveraged. Keep in mind that paralegals and entry-level technical resources may be a good starting point for raising the awareness of predictive coding within a records management program.
- Standards of Care must be addressed in implementing predictive coding. See: Global Aerospace, Inc. v. Landow Aviation, 2012 WL 1431215 (Va. Cir. Ct. April 23, 2012); Moore v. Publicis Groupe, 2012 U.S. Dist. LEXIS 23350, 2012 WL 607412 (S.D.N.Y. 2012); Kleen Products, LLC v. Packaging Corp. of America, No. 10–C5711, Dkt. 412 (N.D. Ill. Sept. 28, 2012).

- Email threading can speed the archiving of information. Email threading software visually groups messages, usually in a hierarchy by topic, with any replies arranged visually near the original message.
- Utilizing some of the rules, you can deploy a solution based on email archiving.
- The chosen tools must always meet the firm's expectations for accuracy and risk tolerance.

OVERALL BENEFITS AND CONSIDERATIONS

Predictive coding appeals to IG professionals because it leverages the efficiencies of TAR tools. This reduces the need to evaluate every document before making a decision as to its business relevancy. It can also provide a platform for identifying electronically stored information by highlighting the primacy of relevant documents, abating the risks related with unreliable analysis and reducing the cost and effort often associated with large-volume classification and filing. When properly utilized, it can surpass the accuracy of human assessment while significantly reducing the volume, and therefore the cost and time, associated with filing and remediation projects.

That said, could predictive coding be an alternative to a DMS? Can it be trusted to meet the challenges of managing content "on demand?" Given its proven ability to identify content through auto-classification with accuracy greater than humans, is it the road to IG nirvana? In a data-saturated environment, where even drag-and-drop technology is considered inefficient, is predictive coding the holy grail of IG? Before answering these questions, let's consider a few things.

It is important to note that predictive coding is not a substitute for IG. It is simply a tool that enhances decision-making capabilities.

In fact, an effective predictive-coding implementation relies on sound IG practices. Training methodologies for the system must first be established to minimize errors and to determine what needs to be classified manually. One cannot do away with human judgment, which is needed to create the learning data set. Predictive coding is only as reliable as the experts who provide the seed documents and ensure the competency of the training procedure. To get the best results and adoption, the IG team should be an active participant in any predictive-coding implementation. Well-designed predictive-coding technologies, coupled with skilled human supervision, increase the rate of identification and classification of relevant documents and considerably improve the chances of success.

Before considering predictive coding technology as an IG tool for your law firm environment, establish consensus on the following:

- Skilled legal experts are responsible for creating the appropriate data sets (or seed sets) and must have access to, or knowledge of, all databases where content resides. The data sets should represent content from all information repositories.
- The product must be able to meet your end-user, IT and legal compliance requirements. Therefore, it is crucial to detail those requirements by defining your business process and use case.
- There needs to be adequate oversight and a comprehensive remediation plan, agreed upon by all stakeholders (Business, Technology, Compliance).
- The deployment should include a process to audit the application's decisions.
- When available, use internal eDiscovery resources to help guide the deployment. Litigation technology experts have been working with this technology for years and can provide valuable insight into its usability and functionalities.
- You do not have to choose between "upstream" or "downstream." A hybrid approach to managing data proliferation is probably the best solution since it will more than likely require different tools.
- Predictive coding is not a panacea, so any project needs to start with the establishment of an IG framework.

There are some arguments against predictive coding as an IG solution. Few law firms are currently using it outside of their eDiscovery platforms; therefore, there aren't many case studies to support the effectiveness of this approach.

There are even fewer standards and documented workflows that can help provide a blueprint for establishing a process and sell the concept to management. Predictive coding and filing tools do not provide a historical set to demonstrate how the results were obtained, as opposed to DMS systems where the resulting information is tracked. Due to regulatory requirements and ethical responsibilities, we must remain mindful of the potential risks associated with revealing sensitive/confidential information.

GETTING TO "GREEN LIGHT:" SELLING INTERNALLY

Needless to say, putting together a comprehensive predictive coding solution will require a commitment of resources and support from firm leadership. Much like many IG initiatives, success hinges on a variety of factors. It is important to develop a strategy that appeals to firm structure and culture. A few suggestions for developing that strategy are to:

- Identify key stakeholders from the business, technology, risk and finance groups.
- Seek buy-in from technology committees, the C-suite, the General Counsel's office, department heads, etc.
- Define the risks associated with unstructured data. Highlight the inability to apply legal holds, ethical walls and disposition to unstructured data.
- Provide statistics on the volume of the unstructured data and the effort required to manage large volumes of electronic information.
- Show the cost associated not only with the storage, but also with IT resources required for the maintenance of sizeable information strata.
- Emphasize the increased systems performance when databases are unencumbered with redundant, obsolete and trivial data.
- Demonstrate the current resources used for the linear review of information.
- Promote lawyer efficiencies associated with their abilities to collaborate with one another and with clients.
- Emphasize the ability to leverage metadata for Knowledge Management.
- Seek partnerships with those already using the technology successfully (eDiscovery, Practice Support, document review teams, etc.).
- Develop a use case and demonstrate how this application can feasibly be applied to an existing problem or issue.

TIPS ON DEVELOPING A PROOF OF CONCEPT

Developing a proof of concept (POC) will give you a chance to test your methodology in a controlled environment and will provide the blueprint for scaling firm-wide. It is important that those involved understand that modifications to existing workflows will be necessary to test the capabilities of this technology. Also, when piloting a predictive technology options, one must carefully consider the impact of reconfiguring a tier-1 application while sensibly mapping a rollback strategy. Make sure you work closely with your application vendor on developing your pilot. Obtain a software map and verify which tools might already have embedded TAR or predictive-coding technologies. If you have a solution in use for eDiscovery purposes, ask for alternative pricing models for use in your IG initiatives. Start by "eating your own dog food" and test your approach with the stakeholders and business unit teams championing this initiative (where appropriate). This will provide intimate knowledge of the predictive-coding framework to those involved with socializing this effort.

While it is necessary to consider all types of persona when piloting new technology, in this case try to identify a pilot group led by technology-savvy practice leaders. Set up a trial instance or partner sample.

There is an ever-growing list of vendors in the market that offer enterprise-wide predictive coding and analytics capabilities. For additional information regarding leading industry vendors, please refer to Gartner Research.

TRAINING THE PREDICTIVE CODING SYSTEM

Much like the concepts we discussed earlier relating to training the predictive-coding system for eDiscovery, the IG plan to use TAR/predictive coding in a strategic manner must be defined and should be the responsibility of the person with IG oversight. The training and setup of the system should be mapped out with proper IT support.

Working with your eDiscovery and Practice/Litigation Support group to understand the resources already available and used in-house can give you a head start toward identifying existing tools for this initiative. Leverage vendor relationships, arrange to participate in their training and updates on software capabilities and uses and learn how other organizations are using these technologies. Establish a list of key words, synonyms, common misspellings and abbreviations. Many firms will already have these lists established and in use in their conflicts of interest departments and/or built within the practice support department's eDiscovery solutions.

Have meetings with subject matter experts, departmental business units and practice areas. Learn how they are working and how they are classifying documents, and understand their data-mapping schematic. In a perfect world, the goal is to always know where all the information is for a specific case file or matter. The reality is that our workforces are made up of talented people who will find creative methods to manage and store data in a manner that is convenient for them, without considering the ramifications of their choices. Security, retention and disposition issues and the need for collaboration are primary concerns for IG professionals, but are often an afterthought — or a "non-thought" — to others. Predictive coding excels in highly structured source information and concept searching, allowing attorneys and staff to work as they wish and still support IG functions, such as data lifecycle management and ease in finding data wherever it may be located.

It is important to validate the approach and methodology for training and using this technology. For example, benchmark a test data set of 500 documents, where 50 documents are reviewed and coded for document setup. The target may be a 95% accuracy rate; once that is achieved and you have built in quality-control checks, the organization can move forward with a successful deployment.

CONCLUSION AND NEXT STEPS FOR FIRMS

The intent of this paper is to continue the conversation on the effective use of predictive-coding technology in an IG strategy. As demonstrated by our discussion of potential areas of use, including unstructured data and RIM functions, predictive coding approaches should be a part of a firm's current and future dialogue relating to governance. The use of TAR does not replace IG, but it does provide another tool that can move firms forward.

The same challenges of using predictive coding for eDiscovery are present in its use for IG. Standards-of-care concerns must be addressed in implementing predictive coding, which includes training the system and verifying accuracy. It is also important for the technology to be evaluated objectively to ensure it is the best tool to accomplish the IG goal. Before implementation, ideal and practical approaches must be addressed. Ask questions, such as "What are we trying to accomplish," and "Is this the best tool to accomplish this goal?" Pulling information from multiple data sets and then loading it into a review tool may require expensive IT resources or a specialist to extract and import data between multiple systems. Firms should evaluate what software and resources they have within their organizations to address the internal needs of IG. Keep in mind that paralegals and entry-level tech staff may be a good entry point for using predictive coding and building strategic partnerships that enhance the firm's IG program.

While these technologies hold promise and potential, it is important to remember that there is a significant commitment required to embark on this type of initiative. Successful implementation requires people and processes to train and retrain TAR systems; education for personnel on how to effectively use the technology; and people to review, validate and audit not only the data results, but the effectiveness of the overall process. With this sort of commitment, there must be some objective measure of success to justify the efforts, such as a reduction in unstructured and untagged data and a reduction in non-essential information. Despite the effort, risks and pitfalls, the potential of predictive-coding technology to support IG is too great to ignore.

Using Information Governance as a Foundation for a 24/7 Law Firm

Compiled by **Brianne Aul**, *Senior Manager, Firm-wide Records, Reed Smith LLP*
and **Charlene Wacenske**, *Senior Manager, Firm-wide Records, Morrison & Foerster LLP.*
Additional contributors are listed at the end of this report.

WITHIN THESE FOUR WALLS

If you have turned on a television even once within the past decade, chances are you stumbled upon one (or twenty) home improvement programs. While the obstacles on each show might be different, the general concept is the same: Current or prospective home owners are seeking to improve their living environments. With assistance (and a little coaxing) from the host, they prioritize their wish list to determine what is a "must have," and what is a "would like, but can live without." Inevitably, additional concerns arise (e.g., cost, location) that force them to review, and perhaps revise, their initial listing. While some concessions might be made on behalf of the owners, the general expectation is that they will be pleased with the end result, and will describe how it ultimately prepares them for future plans.

While law firms have also undergone many changes in the past, these "renovations" have taken on a whole new meaning and approach in the past decade. The "four walls" of a law firm have stretched to include such areas as personal homes, hotel rooms, airport gates and other locations from which an attorney might be found responding to an urgent client matter. As a response, many firms have implemented new technology, processes, and personnel structures to better equip attorneys with the appropriate resources to work – quite literally – anywhere they need, at any time they need. However, this has also opened the door for new processes to be created, procedures to be reviewed, and an overarching need for a solid information governance program.

THE 24/7 "FLOOR PLAN"

Telecommuting is on the rise in many industries, and carries with it several benefits: the time savings on commuting, the reduction in costs associated with working from an office, and the convenience and comfort of working in a more familiar environment. However, there are other reasons why many larger law firms have migrated to the "24/7" structure. While some firms remain decentralized in their infrastructure, others have elected to use certain systems across the enterprise, and thus, they might require support from a more centralized team. As clients also have more of a global presence, the opportunity for their questions and needs to arise around the clock has also increased. And, of course, attorneys have become increasingly mobile – it is commonplace for an attorney to be working on multiple matters at once, en route to a client location while feverishly addressing another client request.

While there may be differences in terms of how firms implement a 24/7 business model, it's generally assumed that they demonstrate one or more of the following:

- Firm members (or certain firm members) have the ability to remotely connect to the Firm's network.
- The firm has a mobile device management program in place.
- The firm has offices across multiple time zones.

- The firm has clients across many time zones.
- The firm's staffing structure spans several shifts.
- The firm has administrative teams with global responsibilities and/or has a consolidated staffing center for global requirements.

While 24/7 firms operate predominantly on access to electronic data, in situations where physical records are still needed, there are typically minimal restrictions on transporting files outside the building. However, some 24/7 firms are investigating the imaging of physical records as an opportunity to provide access to records from remote locations, as well.

CRACKS IN THE FOUNDATION

Most home improvement shows emphasize on the importance of careful inspections. Likewise, it is crucial for law firms to look closely and identify flaws in the 24/7 model, which may include the following:

CONSUMER TECHNOLOGIES

As personal laptops and tablets tend to fall out of the firm's domain, they have a higher exposure to security concerns. Their owners may not have a password enabled on their personal computers, or the personal passwords may not be subject to standard "strength" requirements (e.g., special characters, alphanumeric requirement, etc.). Additionally, users may elect not to use anti-virus or encryption software on their personal computers, or may not be consistent in upgrading the software as needed. Users might also save materials on home drives or upload/download material on unencrypted USB drives, which again, falls outside of the firm's ethical wall systems or other information security measures.

PERSONAL EMAIL, EXTERNAL FIRM REPOSITORIES

In certain situations, users might elect to email and access client documents via their personal accounts. This is problematic because personal email is not subject to retention policies that the firm dictates, presenting an opportunity for information to be shared outside of firm-governed collaboration or communication tools. Additionally, transmission of client information via personal email can be less secure than information sent through a firm's email system.

In similar fashion, users might use external file hosting sites to share information, which, again, falls out of firm security and document tracking mechanisms. Depending upon the site used, there may also be additional concerns regarding the site's own security and risk for hackers, etc. Additionally, when external file-hosting sites are used, the question of data ownership arises. Firms run the risk of data being shared or retained by those hosting the site, unless appropriate safeguards and agreements are in place.

MOBILE DEVICES AND BYOD

Mobile devices, whether personally owned or provided by the firm, can be lost or stolen. These issues are not always reported to the appropriate individuals, and thus, there is an increased likelihood of client or firm data being accessed by others — especially if the mobile device has not been encrypted. Additionally, with mobile devices, the line between what is firm (or client) property and what is personal property can become muddled without a defined policy and process. This is especially prevalent in cases of departing individuals who use their own mobile devices.

PHYSICAL RECORDS

To the extent that physical records are utilized, they are subject to the care and attention that the user provides. If firm personnel do not notify the appropriate department that they will be removing the file from the firm, tracking of the record becomes increasingly difficult. General maintenance of the record is also a concern. Physical files may be lost or stolen during transit from office to home, and if they are not on appropriate media, the contents could be subject to damage from external factors. To the extent that content would ever be intentionally destroyed outside of firm premises, firm personnel that fail to use a proper method of shredding, or who do not even have access to a shredding method, pose a threat to information security.

ENVIRONMENTAL CONCERNS

The ability to work from "anywhere" means that certain environmental factors can impact information security. For example, users can connect to unsecure networks in order to access firm information. A personal device used at home might also be shared by other family members, which could result in data being inadvertently accessed or compromised. In more public settings, such as coffee shops, conversations can be overheard, or screens can be viewed — and shared — by individuals.

DATA PRIVACY

Since "anywhere" can also mean any country or any region — and because certain privacy laws, or even client requirements, might prohibit data from being accessed from certain regions — there is a concern with data being accessed remotely by firm users. A user unaware of the data privacy laws could inadvertently view information that otherwise should be restricted.

DEVELOPING THE BLUEPRINT

By understanding the workarounds users may have developed and proactively addressing them, the firm should be able to mitigate a large portion of the associated risk operating "24/7." However, there are crucial pieces that need to be in place, including policy, procedure, and education.

POLICY

Established policies regarding how data should be handled and what all personnel are expected to do to handle firm and client data should be established and communicated — frequently and consistently — to all members of the firm. It is important to ensure that the policies are aligned with both firm culture and the firm's level of risk tolerance. It is also crucial to ensure that these policies do not remain stagnant; they should be reviewed and refreshed continually as technology changes.

While not a comprehensive list, such policies should focus on the following:

- Data should only be placed and shared in firm–approved repositories, and located within the appropriate designated areas, which fall under firm security and appropriate matter membership.
- All potential breaches of data should be promptly reported to firm management.
- Any loss of firm–provided mobile devices or personal mobile devices containing firm data should also be reported promptly to the designated department within the firm.
- Appropriate methods of client communication should be clearly defined (e.g., establishing policies on social media, instant messaging, etc.).
- Personal devices that might be used to access firm and client data should be required to be password protected and have anti–virus protection enabled.
- Client and firm documents, including emails, should not be permitted to be sent to home email accounts.
- In the event of a departure, data related to the firm and its clients are firm property and should be wiped from a personal mobile device accordingly.
- Data that is restricted by ethical walls or other security measures must be addressed and approved by the appropriate team prior to being provided to a requesting user, if the user is not already permitted access.
- Any physical records being taken from the building by timekeepers should be tracked, secured, and returned to the firm. In some situations, it might be required that certain records with higher security requirements should not permitted to be removed at all.

PROCEDURES

In connection with developed policies, there should also be established procedures for users to follow regarding data being used outside of firm offices. Although the procedures should be outlined with the established policies, there is also a need to make them "simple" to follow, while still achieving the needs of the policy. A firm should ensure that procedures are established for such items as:

- How to determine whether a user should be granted remote access to the firm's network.
- How, and to whom, a user should report a data breach or loss/theft of a device containing firm data.
- How to seek approval to provide a remote user with access to an otherwise restricted document when the request is urgent and/or off-hours.
- How, and whom, to notify if a physical record is being removed from the building.
- How to return or destroy data in the event of a departure from the firm.
- How to seek approval for utilizing a client-preferred data-sharing site (this should be subject to rigorous testing and education of the risks, to the extent that they exist). It may be that certain requests are never approved. To the extent that they are, there should be a sign-off procedure from both attorney and client — if they elect to move forward — that they are aware of, and accept, the associated risks.

EDUCATION

Policies and procedures are crucial to risk prevention in a 24/7 firm, but education is equally vital. Most users will perform the appropriate procedure, or follow the policy, if they not only a.) are aware of it, but also b.) understand the reasoning behind it. Users should be aware of the ramifications of accessing data outside of firm protocol and the risks it presents to not just the firm, but clients as well. Additionally, users should be aware of any legal or client requirements that the firm must follow, so that they can act within the established boundaries. Understanding the impact on ethical obligations and client relationships — more so, the detriment to both when policies are not followed — helps highlight what the firm is attempting to promote or prevent.

Education should be thorough, and remain current with any new firm processes or technologies. It should be a requirement as opposed to a suggestion, and should also be consistent with what the firm does. As often happens, education works best when verbal instruction is combined with action. By demonstrating that those in senior management are already adhering to the process, it creates and serves as an example for others to follow.

STRUCTURAL SUPPORT

While in many situations the onus falls upon the individual user to make the 24/7 firm successful, there are additional opportunities for firms to help facilitate the "around the clock" structure. As mentioned above, the 24/7 firm likely has certain departments working multiple shifts to accommodate requests from remote users or other offices. However, those departments may be restricted to only certain groups, such as IT. Metrics should be used to determine what other types of requests are arriving throughout a 24/7 period. In evaluation of those metrics, it may be that the firm determines other departments should also have multiple shifts or that certain "power users" be "on-call" for any questions that arise during a given time period.

Additionally, business continuity initiatives, such as planned system maintenance/outages, should be prescheduled and clearly communicated to users, so that they may plan accordingly. In situations where an outage will be longer than the standard, or will impact a more critical system, considerations should be made as to what individuals can use in the interim. In addition to certain data privacy requirements, it may be that multiple instances of certain software exist in the firm's infrastructure, so users in one region are not impacted by an outage planned in another.

Images of physical records carry a variety of benefits, but as part of a 24/7 firm, they also have the potential to be accessed in lieu of the physical record itself. So long as the image is classified within the firm's appropriate data repositories, the potential for physical documents to be lost or compromised is significantly reduced. Imaging also enables collaboration amongst teams by providing access to the data to multiple members — and allows for users to easily access the information without the requirement of transporting the physical record itself.

UNDER CONSTRUCTION

While the 24/7 attorney has been in existence for several years, the 24/7 firm remains a work in progress, as new technologies and staffing structures are developed to support it — and new security measures are implemented to protect it. As firms continue to prioritize cutting-edge technologies and processes in order to provide clients with the highest levels of service, it is clear that the need for a 24/7 model will only increase. And as it does, it will demand new information governance measures over time. These measures will help ensure that the 24/7 firm not only survives, but also thrives, enabling it to remain ahead of the curve in the legal industry.

THE FOLLOWING 2013 SYMPOSIUM PARTICIPANTS CONTRIBUTED TO THE THREE SECTIONS OF THIS *EMERGING TRENDS* REPORT.

Derick Arthur
Director of Records & Facilities,
Proskauer Rose LLP

Odell Bryant
Director of Records and Conflicts Administration,
Cravath, Swaine & Moore LLP

Carolyn Casey, Esq.
Senior Manager Legal Vertical,
Iron Mountain

Beth Chiaiese, CRM, MLIS
Director of Professional Responsibility & Compliance,
Foley & Lardner LLP

Scott Christensen
Director of Technology – US,
Edwards Wildman Palmer LLP

Terrence J. Coan, CRM
Senior Director,
HBR Consulting LLP

Galina Datskovsky, CRM, PhD
Principal, Independent Business and Information Governance Consultant

Esther Diamond
Records Manager,
Locke Lord LLP

Beth Faircloth
Director of Risk Management,
Seyfarth Shaw LLP

Patricia A. Fitzpatrick, CPA
Director of Practice Management,
Katten Muchin Rosenman LLP

Grant W. James, CRM
Firm Records Manager,
Troutman Sanders LLP

Matt Kivlin
Director of Product Management, Legal,
Iron Mountain

Norma Knudson
Director of Office Administration & Compliance Support,
Faegre Baker Daniels LLP

Frank LaSorsa, CRM
Director of Records and Information,
Kelley Drye & Warren LLP

Brian Lynch
Director Risk Practice,
IntApp

Faron Lyons
Open Text Inc

Brian B. McCauley, CRM
Director of Information Governance,
McDermott Will & Emery LLP

Dana C. Moore
Information Governance Compliance Manager,
Foley & Lardner LLP

Eric Mosca, CRM
Director of Operations,
InOutsource

Deb Rifenbark, CRM
Chief Records Officer,
Stinson Morrison Hecker LLP

Steven Shock
Chief Technology Officer,
Irell & Manella LLP

David B. Steward, CRM
Husch Blackwell LLP

ABOUT IRON MOUNTAIN. Iron Mountain Incorporated (NYSE: IRM) provides information management services that help organizations lower the costs, risks and inefficiencies of managing their physical and digital data. Founded in 1951, Iron Mountain manages billions of information assets, including backup and archival data, electronic records, document imaging, business records, secure shredding, and more, for organizations around the world. Visit the company website at www.ironmountain.com for more information.