eDiscovery Roundtable:

The State of Information Governance and Advanced Analytics

MINDSEYE

Content Analyst® Company
Better Analytics for Smarter Solutions

TABLE OF CONTENTS

INTRODUCTION

In today's rapid shift to digital information, the unstructured data created, captured and stored is now a problem for many companies. Some of this data is wasteful, some is extremely valuable, and some carries significant risk. Unstructured data is a universal component of business, but the perceived value and key characteristics of this data are truly in the eyes of the beholder. Risk — relating to compliance, legal, human resources, and other areas — spans across all business units of an organization[1]. However, risks are interpreted differently based on the business function, and risk doesn't necessarily boost value or profits — it creates concern and non-action.

In the era of big data, companies are beginning to see the significant value that can be captured by evaluating trends hidden in mountains of unstructured data. The key to making sense of data, however, is in understanding what unstructured data you have, eliminating what isn't needed, and defining what information is key for the business — in reducing risk and increasing value. In the 1800s, the untrained eye saw the Sierra Nevada as a mountain chain, but those that knew how and where to look saw the precious gold mine that lay just beneath the surface. The same can be said for data today.

Innovations in technology are reshaping the ways in which unstructured data can be mined, understood, used, and managed. Tighter regulatory and security concerns are driving organizations to explore better ways of managing their unstructured data. Data touches every aspect of a business, and the processes for managing it can be much more complex for supporting multiple stakeholders. There's never been a clearer need for renewed focus on information governance (IG) in the enterprise and more advanced ways of deriving intelligence from the process. This was the topic of a recent roundtable discussion hosted by Mindseye and Content Analyst Company, in which participants shared their experiences in information governance, how they are handling the growing volumes of unstructured data, where IG could be improved or implemented more effectively, and the role advanced analytics can provide to the process.

The participants of the roundtable were assembled from Fortune 500 corporations and Am Law 250 firms from North America, each with unique experiences and perspectives on the topic. The participants came from varying backgrounds including leaders in legal, IT, records management, knowledge management, litigation support, and consulting. This report summarizes the participants' perspectives on the state of IG, the challenges faced in defining the process, the organizational barriers of implementation, and the role of people and advanced technology in ongoing improvement.

[1] A practical guide to risk assessment, PWC (December 2008)

UNSTRUCTURED CORPORATE DATA: MANAGING VOLUME AND COMPLEXITY

Discussions among the roundtable participants confirmed that the volume of unstructured data is an ongoing and growing challenge. It's easier to manage this growth with inexpensive physical storage than it is to develop, document, and enforce policy, which requires political strength in the organization. Unlike storage, management requires addressing the overall complexity of classifying unstructured data for the specific needs of each business unit while, at the same time, satisfying legal and compliance obligations.

The big story about unstructured data isn't just growing volumes.

For the past few years, the big story about unstructured data has been its explosive rate of growth, where 90 percent of the world's data has been created in the past two years[2]. But the roundtable participants identified another critical issue, which is the explosion of data that is in non-traditional formats. This goes beyond social media (which one participant said "makes attorneys want to put their fingers in their eyes") and rogue clouds (which "keep IT and legal up at night"). There's a wide range of new formats that contain valuable, or potentially risky, data that aren't as easy to manage and find as traditional types of data. One participant cited an example with heart rate monitors, which store information that could potentially be used for litigation. Moreover, images are forming an increasing proportion of potentially relevant information. Another participant cited the banking industry, where transactions are increasingly being recorded with check images. People are using images to communicate in ways they never did before — through pictures.

Storage has significant hidden costs and complexity.

Advancements in technology have continued to drive the cost of physical storage down; however, the cost of managing the growth of unstructured data is far more significant when factoring in the overall costs. Some of the hidden costs identified by participants included:

[2] Big Data, for better or worse: 90% of world's data generated over last two years. ScienceDaily (May 22, 2013)

- » Productivity lost to trying to locate files, records and emails, or to recreating data that already exists

- » Legal and compliance costs associated with the storage of duplicative or near-duplicative content

- » Ongoing management of the IT systems and backups associated with the data, environmental costs associated with running the hardware and data centers used to house unstructured data, and the inherent cost of both risk and exposure of sensitive material — which is immeasurable

Adding to the complexity are the ever-increasing ways that unstructured data can be stored. For example, many companies have had policies for — and have been blocking the use of — USB drives for several years. But now users have many other storage mediums at their disposal, including iPads, personal clouds such as DropBox and Google Drive, smart phones with corporate email accounts, etc. Known as "bring your own device" or BYOD, this phenomenon poses IG challenges[3] and current policies may be inadequate to address the use of these other types of storage. For example, one participant mentioned that an organization may have policies that define what an email or a business document is, but not where people are allowed to save it — local machine, mapped drive to the server, disk, thumb drive, home computer, DropBox, Google Drive, etc.

Discovery, data security, and retention are the issues.

In the course of the discovery process, documents, records and emails may be captured on a hard drive, sent to a law firm, and then forwarded to a service provider. Many law firms do not proactively manage retention of client data. Participants were asked how they validate that everyone is compliant with their organization's security standards. Participants acknowledged that many firms simply cannot confirm that all of their client's data has been purged from the firm's systems at the conclusion of a matter, let alone from a selected service provider's systems. Participants felt strongly that corporations should ask their firms for data retention schedules, as well as a list of all vendors and anyone else with access to their data and their retention schedules. Policies are important to address today's data challenge. But even well-crafted policies are meaningless without ownership and enforcement. And that's where IG comes in.

---

[3] Information Governance in the Age of BYOD, by Jason Glass, Information Discovery Digest (November 13, 2012)

Key conclusions

- » It's not just the volume and growth (velocity) of data that create challenges. Data variety, such as BYOD and different unstructured data types, makes it difficult to manage.

- » The cost of storage is not the issue, rather it's the cost of finding relevant documents for litigation that has greater potential impact for the organization.

- » Better security practices and controls are needed to ensure that data used externally by law firms and service providers in litigation gets promptly destroyed at the conclusion of the matter.

WHO OWNS IG? THAT'S A GOOD QUESTION.

Participants mentioned that they are beginning to see organizations making a conscious effort to take a business-oriented approach to information management. Many participants underscored the inherent challenge with IG, which is that the process needs an owner with the political clout within the organization to drive change. This is accomplished through effective communication of value to each stakeholder so they understand the importance, a well-defined policy, and an ability to enforce the policy to change behavior. The owner must also have a great supporting cast of stakeholders who have input into the process, policy and benefits. These stakeholders must have a unified message that communicates the value of the process to ensure wide-scale adoption.

Information must be viewed as an asset — and business ownership is critical.

Participants agreed that in order for corporations to prioritize information management, they need to view their information as an asset, which may require a change in perception. Many companies consider information as an asset when it uncovers key patterns that help increase profitability, whereas managing risk can potentially be viewed as a cost to an organization. Sensitivity to risk will vary based on the industry and business type, and depends on how the organization views its assets. For instance, a manufacturer will have a different perspective of the requirements and value of its data and data sources than a financial firm will.

Today, information has become a significant asset[4]—and a high risk—but it may not be managed that way. This requires business ownership to change organizational thinking, and many participants agreed this is currently lacking. Participants did, however, mention that some companies are beginning to create a new executive role: the chief data officer[5]. The CDO plays a different role than the CIO, but the two must collaborate closely. One participant put it this way: "The CIO takes care of the pipes while the CDO focuses on the water that runs through the pipes."

Organizations may not be asking the right questions.

The consensus seemed to be that everyone wants IG, but the challenge is in figuring out how to get there and where to begin. The key question that needs to be asked is: "What is the organization trying to solve and how will we measure success?" Participants acknowledged that often, organizations don't have a clear understanding of the overall objectives and that objectives vary by business unit. Nevertheless, they rush to implement technology before defining the goals of the organization and the needs of each stakeholder—which ultimately ends up in failure. They buy the technology without testing it because they don't know what to test it for. One participant recounted his own company's experience implementing a "solve-all" technology that ultimately needed to be scrapped. This is, unfortunately, not a rare occurrence and it can have a devastating impact on future initiatives.

Key conclusions

» A natural lack of ownership of unstructured corporate data has led to the development of "chief data officers."

» A holistic approach needs to be taken to address information governance issues.

» Goals and objectives need to be clearly identified and communicated before selecting appropriate technology to address the needs.

TECHNOLOGY

Participants acknowledged that their businesses do actually view some information as an asset, and their organizations build and buy the best, most appropriate technologies for capturing as much value out of this asset as possible. For instance, the financial services industry has been using technologies around predictive analytics for years and has some of the most sophisticated solutions for understanding data that helps drive investments. However, when it comes to managing business data for risk purposes, the technology solutions purchased often fail. The differences between the two technologies may not be all that different, but the research, problem and objective definition, and organizational alignment drive one focused solution to succeed while another flounders.

Technology is neutral and technology alone is not the answer.

Technology itself is neither good nor bad. One participant summarized it as simply a magnifier of good or bad, meaning if you automate a poorly defined process, you can do more harm, much quicker.

Sometimes organizations need to solve a problem quickly and they hastily apply technology without enough understanding or planning. But to deliver the promised benefits, technology must be implemented and supported by the right people to automate the right processes and enforce the right policies. It's more than just people and processes—it's having the *right* ones. Knowing this requires a complete understanding of the problem, but organizations often try to solve problems without taking the time to truly understand them. Then, when the "solution" doesn't work, it affects the ability to address bigger problems within the organization.

Selling technology within an organization can be difficult.

On the one hand, technology can be politically easier to sell than people and process. You can buy a specific widget to meet a specific business need. But because you need the process and the people engaged with the chosen technology, it can be hard to get executive buy-in on the "whole package"—until there is an emergency or business crisis, or litigation arises.

The primary components that go into making any purchasing decision are business benefit, timing, budget, and politics. In many organizations, budgeting may

[4] Invest in Information and Analytics to Benefit From Big Data, by Douglas Laney and Frank Buytendijk, Gartner Research (March 8, 2013)

[5] Chief Data Officers: Culture Change a High Data Management Hurdle, by Ian B. Murphy, Data Informed (July 17, 2013)

be done on different timelines. For example, publicly traded companies report on a quarterly basis, but some departments may make annual budgets and other business units focus on the longer term. In addition to navigating multiple cycles, each business unit will look at the initiative and ask, "Is it good for me?" to determine the business benefit — and the answer may change from division to division. Even when the timing is right and the budget is available, it can be difficult to get through the political silos. Everyone in an organization needs — and creates — unstructured data. In fact, it's one of the few things that touches every facet of an organization. Each business unit has different ways of viewing and using it with varying perceptions of the inherent value and risk. How do you build a business process that spans multiple business units when each stakeholder has a different use case for the data and different thoughts on what and how technology should be applied, if at all?

Key conclusions

» Wide-ranging initiatives like IG require technologies that can address the various needs of the stakeholders. Often, there is no single solution that will achieve the goal.

» Technology alone cannot address IG issues — a deeper understanding of the problem, combined with the right processes, is required to address these challenges.

» Without key ownership and communication of value, technology initiatives may not get required support and buy-in, which can slow adoption.

ADVANCED ANALYTICS

Successful IG initiatives leverage technology to classify information[6]. Classification isn't a new concept — in fact, most people use auto-classification every day. Email is a great example: spam filters help classify junk versus non-junk, and many people use rules to organize their incoming mail[7] based on metadata and keywords rather than manually moving each individual message one at a time.

Organizations aren't capturing long-term value on the unstructured data that's collected for specific legal matters.

Law firms participating in the roundtable admitted that they haven't fully embraced auto-classification and other advanced analytics technologies today for use in legal matters. Law firms are typically hired for a specific legal event, not for ongoing IG consulting or policy development and enforcement. Because of this, firms are not focused on early classification until litigation arises. By that time, firms should be well-versed on their corporate client's policies, procedures and enforcement of IG initiatives. They can leverage these during initial discovery efforts and understand where the data resides, scope the volume, and develop a case strategy complete with budget.

Taking this a step further, law firms can leverage advanced analytics when receiving data that has been collected from their corporate client. The use is two-fold: one, to be able to initially classify the data collected and create a streamlined review strategy by applying other analytics attributes; and two, to apply advanced analytics to the same data on other or future litigations. Participants agreed that it makes no sense to collect the same data more than once, incurring repeated expenses, when analytics can be applied to classify and focus the needed data on other matters. To benefit from significant time and cost savings, IG-focused organizations need to either demand these capabilities from the firms that they hire or contract a consultant or service provider to assist them.

Finally, it may be beneficial for an organization to request a post-matter debrief. Most of the participating firms do not offer these without being asked. A post-mortem that evaluates what was done with the collected data and what could be done differently to increase the rate of review would enable the business to capture value on the data they've collected that goes beyond the specific case. It also would provide the firm with an understanding of how to better serve this client on future matters. This requires firms to take a client-centric, rather than a matter-centric view.

Take a "weed, organize, and refine" approach when applying advanced analytics to IG or legal matters.

When it comes to leveraging advanced analytics for IG, the initial approach might be the same as it would be for a legal matter. Get rid of the weeds first! In litigation, rather than

---

[6] The Value of Categorized Documents, by John Felahi and Trevor J. Morgan (2013)
[7] Use Rules to Automatically Organize Email, by Zack Stern, PCWorld (December 23, 2008)

trying to find the smoking gun right off the bat, the first focus should be on weeding out the irrelevant information[8]. In either scenario, this helps to narrow down the amount of data that needs to be combed through or analyzed more closely. This allows for segregating the old information from the new, the personal from the non-personal, the responsive from the non-responsive and many other types of information classification.

Participants recognized that it can be a challenge to standardize on the classifications to be used because most organizations try to get too granular rather than focusing on larger categories. In reality, organizations can build the classification — even at a high level — then "auto-categorize" new data accordingly at the time of creation, using advanced analytics that are concept-aware. From there, they can refine the categories over time. By taking a weed-organize-refine approach, categorization is fast and efficient, and what is left is QC.

Key conclusions

» IG-oriented organizations should leverage advanced analytics to implement their IG policies and procedures.

» Advanced analytics can reduce costs when applied to data on receipt, saving time in the review process and organizing the data for future use.

» Using advanced analytics to categorize unstructured data can help reduce the volume of data not just in legal matters, but across the broader landscape of information governance.

FINAL THOUGHTS

The participants in this roundtable discussion agreed that solving problems before better understanding them leads to initiatives that attempt to "boil the ocean" — which simply can't deliver a successful outcome. A better approach is to break down the ocean of seemingly insurmountable IG challenges into more manageable problems that can be defined and solved. Ensure that due diligence is taken when addressing the long-term issues, when evaluating the technology that would be most beneficial to address those issues, and identifying which people and procedures will be deployed to ensure success. Select an area where positive change can be effected such that if the project fails,

it becomes a learning experience without having a negative impact on future initiatives. Instead of solving for the enterprise, solving for the business unit and then expanding from there may yield the best results. Or, if the goal is three years out, begin by breaking it down into smaller milestones, which can help get everyone on board and keep them engaged. Early trial and error can help identify what changes need to be made quickly and how to evolve from there to build policies, uncover classifications, and more. By leveraging advanced analytics throughout the process, organizations can better classify and understand legacy data while minimizing the burden on end users and records management professionals going forward.

FOR MORE INFORMATION, PLEASE CONTACT:

Mindseye
sales@mindseyesolutions.com
(888) 770-3876
www.mindseyesolutions.com

Mindseye is a leading provider of eDiscovery software solutions. The company's flagship eDiscovery platform helps organizations manage risk, minimize legal exposure, and eliminate wasted time and money throughout the discovery process. Delivered through a fixed-cost licensing model or as a stand-alone eDiscovery Appliance, Mindseye is a scalable solution built on an iterative workflow design. Organizations that use Mindseye can rapidly process and use early insights gleaned from the data to make cost and resource estimates, formulate strategy and case direction, and ultimately move less but more relevant data to review.

Content Analyst Company
info@contentanalyst.com
(888) 349-9442
www.contentanalyst.com

We provide powerful and proven Advanced Analytics that exponentially reduce the time needed to discern relevant information from unstructured data. CAAT, our dynamic suite of text analytics technologies, delivers significant value wherever knowledge workers need to extract insights from large amounts of unstructured data. Our capabilities are easily integrated into any software solution, and our support strategy for our partners is second to none.

---

[8] The Future of Attorney Review: Smaller, More Relevant Data Sets, by Jeff Fehrman, Focused Discovery (August 15, 2013)