Predictive Information Governance: The Time Has Come

June 13, 2013

By: Barry Murphy

Table of Contents

You Have Only Heard A Fraction Of The Big Data Story

Big Data has become big business. The ever-rising volumes of digital data in the global corporate world create opportunities for companies to leverage information in ways never before possible. Leveraging information requires the ability to quickly and efficiently analyze it and understand what it means and how it can affect business decisions. Examples abound of innovative uses of analytics to mine Big Data:

- A financial services company reducing the amount of time necessary to run market surveillance algorithms so that compliance officers can quickly spot questionable stock trading trends.
- A telecommunications provider quickly being able to combine data from a variety of sources—social media sites, call records, cell tower logs, etc. – to see which customers are experiencing a issues such as dropped calls at any point in time and proactively contact the customer and deal with the problem.
- A consumer products company creating granularly targeted marketing plans by forecasting how customers are going to respond in the future based on their past behaviors and their segmented demographics, as tracked and stored in a CRM system.

While these examples showcase the value of leveraging information to gain revenue and cuts, they only hit the tip of the iceberg. That is because Big Data to date has been focused almost solely on structured data – that which fits into the rows and columns of a database and is relatively simple to manipulate for the purposes of analysis. But, even conservative estimates are that structured data makes up no more than 15%-20% of overall corporate information. In other words, companies fail to leverage 80%-90% of the information stored in various repositories. This other information – unstructured content – does not fit easily into a traditional database; rather, content is typically stored in some kind of a file system or repository and the metadata about the file is stored in a relational database.

As a result, running analytics on unstructured content can be more complex than doing so on structured data. That is why Big Data has a lopsided focus on structured data. But, organizations are beginning to focus more on unstructured content and gaining valuable insight to what is in it in order to manage risk, cut costs, and increase revenue.

Information Governance Has Become A Rising Challenge

Organizations tend to put more discipline around managing structured data for business intelligence (BI) because it is a more straightforward activity. There are plenty of good BI tools that can give companies deep insight into ERP, CRM, and other structured systems; the insight gained allows for timely communications with customers, more targeted marketing campaigns, and more efficient customer service among myriad other benefits.

It is easy to turn a blind eye to unstructured content, though, partially due to the fact that there is simply so much of it. Companies create an overwhelming amount of documents, emails, instant messages, and social media content, and McKinsey Global Institute, projects data to grow at 40% per year.[1] Without information governance (IG), the sheer volume of information stands to drown organizations in high costs and myriad risks. Do companies take IG seriously? The data suggests that they are beginning to. While only one-third of companies eDJ surveyed last fall have an active IG program, another one-third have a program in the planning stages.

---

[1] Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Authors: James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers. May 2011. (http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation)
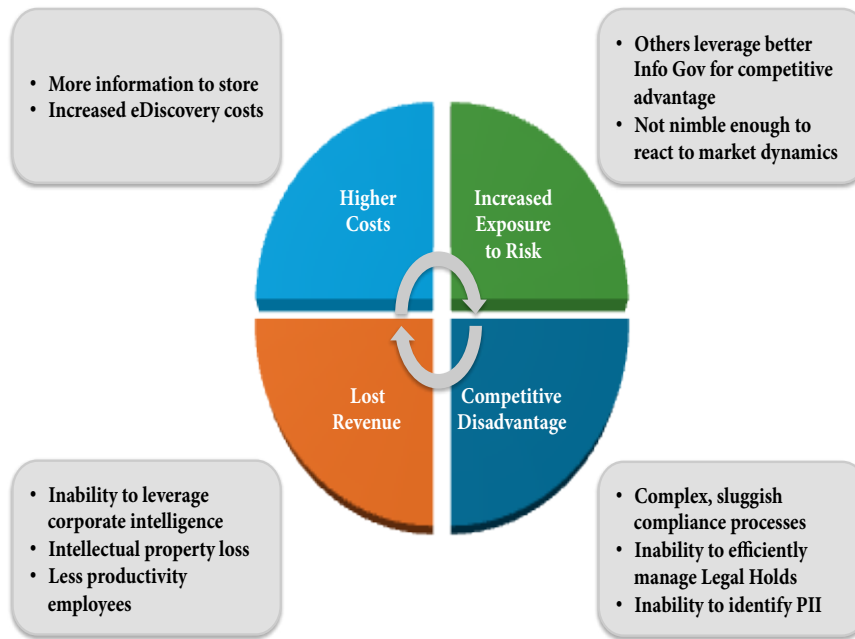
Information Governance Programs Slowly Gaining Footholds In Organizations

Does your organization currently have IG program?

Yes, we have an active IG program	We have an IG program in the planning stages	No	Don't know
33.3%	32.5%	29.9%	4.3%

Source: eDJ Group and ViaLumina Spring 2012 Information Governance Survey N = 117

Part of the challenge in establishing an IG program is the sheer breadth of the topic. eDJ survey respondents consistently see IG as a very broad concept, incorporating multiple disciplines including eDiscovery, records management, security, compliance, storage and archiving, risk management, and privacy. For most, it is overwhelming to think about getting all information under control across so many disciplines. Not addressing the management and governance of unstructured content, though, can have dire consequences.

*e*DiscoveryJournal

The Negative Consequences Of Poor Or Non-Existent Information Governance

- More information to store
- Increased eDiscovery costs

- Others leverage better Info Gov for competitive advantage
- Not nimble enough to react to market dynamics

Higher Costs

Increased Exposure to Risk

Lost Revenue

Competitive Disadvantage

- Inability to leverage corporate intelligence
- Intellectual property loss
- Less productivity employees

- Complex, sluggish compliance processes
- Inability to efficiently manage Legal Holds
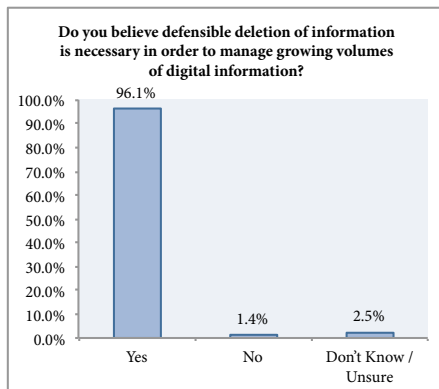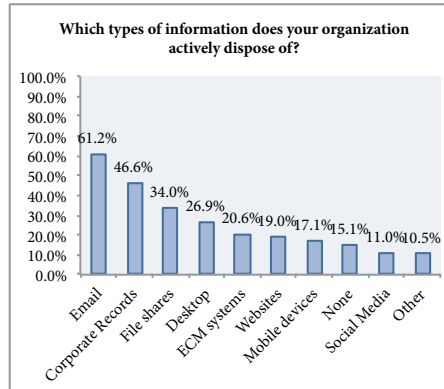- Inability to identify PII

*e*Discovery*Journal*

Some of the inertia around IG stems from the huge gap that exists from the current state of IG to a more desirable state where information assets are easily findable, deleted when their value life-cycle runs out, locked down when they need to be preserved, and protected from unauthorized loss or access. This gap is well illustrated by the huge divide between those that believe defensible deletion is necessary and those that actually delete information.

Huge Gap Between Those That Want To Delete Information And Those That Do

Virtually Everyone Wants To Begin Defensibly Deleting Information

But, Few Delete Anything More Than Email

Do you believe defensible deletion of information is necessary in order to manage growing volumes of digital information?

Yes 96.1%
No 1.4%
Don't Know / Unsure 2.5%

Which types of information does your organization actively dispose of?

Email 61.2%
Corporate Records 46.6%
File shares 34.0%
Desktop 26.9%
ECM systems 20.6%
Websites 19.0%
Mobile devices 17.1%
None 15.1%
Social Media 11.0%
Other 10.5%

*e*Discovery*Journal*

*e*Discovery*Journal*

Until recently, few tools have existed to efficiently extract information from unstructured content and give companies the ability to understand what exists, how to apply policy to it, how to apply automated rules to it, and do all of this in a defensible manner.

Such technology does now exist: predictive information governance. Companies need to understand how to harness this technology, marry it with human expertise, apply it within business processes, and validate what results from its usage.

Predictive Information Governance can be a catalyst for several very important IG projects:

Predictive Information Governance A Catalyst For All Sorts Of IG Projects

*e*Discovery*Journal*

- **Compliance**. Due to regulatory and specific industry rules, it's crucial to ensure that business policies are in place and followed. Most compliance processes, though, tend to be complex and confusing with information flowing through companies at a rapid velocity and in large volumes. By applying automatic categorization to content it is possible to begin automating the execution of business policies to enterprise information.
- **Security & Privacy**. As information volumes grow, it becomes more difficult to make sure that sensitive information does not fall into the wrong hands or that intellectual property does not leave the four walls of the company. There are often strict compliance rules for securing, and even hiding personally identifiable information (PII). Predictive information governance can be used to identify, secure, and redact sensitive information within content (e.g. Social Security Numbers, Account Numbers, Addresses, Names).

- **Information Migration**. As the survey data noted earlier, many organizations desire to clean-up shared drives, often by migrating information to managed repositories such as SharePoint. The challenge with this migration is taking moving content with little or no organization into a system with very granular organization and categorization. There is a need to apply predictive information governance to bring automation and intelligence to the categorization process, as part of the migration. Human intelligence will be critical to the process in terms to utilizing the insight provided by predictive information governance, but the technology can really get these projects moving in the right direction.

- **Defensible Deletion**. As previously noted, virtually everyone wants to defensibly delete information to manage storage costs and reduce risk. The challenge has been ensuring that only the correct information is deleted within mountains of information. It becomes impossible for human beings to check each and every file before its deleted. With the application of predictive information governance combined with human oversight and quality assurance methodologies, this process can be automated and defensible.

- **Records Identification**. Companies have practiced records management for years. But, earlier paradigms – storing paper records, for example – were more straightforward and manageable. In the early years of the digital age, companies struggled to classify electronic business documents against taxonomies for retention, disposition and policy purposes. However, more often than not these initiatives have failed. These failures are typically the result of the inability of business users to consistently and manually classify their documents while keeping up with the volume of information created and received. The use of predictive information governance to augment human classification can bring automation and more consistency to the process. This ensures the correct information is kept and available for the business to use and mine for additional value.

- **Legal Hold**. Companies have huge volumes of stored information and want to begin defensibly deleting that which is no longer necessary. But, some of that information may be subject to Legal Hold. That thought alone can stop some in their tracks and prevent any deletion. Predictive information governance can help to identify content that is likely to be subject to Legal Hold and allow companies to make decisions on that content and preserve it before it can be deleted. When combined with quality assurance measures, such an effort could be a reasonable attempt to ensure that information is preserved.

- **eDiscovery**. Companies need to find and review information for several reasons: internal investigations; litigation; and regulatory requests. No matter the reason, it is necessary to quickly and efficiently preserve potentially relevant data. Predictive information governance can help eDiscovery in several ways. First, they can make collection and preservation efforts more targeted, faster, and well-informed. Later in the eDiscovery life-cycle, Predictive Coding can make Legal Review faster, more efficient, and significantly less expensive than traditional methods.

- **Knowledge management**. In today's world, companies must leverage internal corporate knowledge to stay competitive. Knowledge Management programs of the past got a bad reputation for not providing real value; that is because most of those early programs were focused solely on information access – business users pulling information to themselves. With predictive information governance, companies can begin to push the right information to knowledge workers to make them more efficient and effective.

- **Business Optimization**. When Knowledge Management is effective, companies leverage corporate intelligence to make better decisions, recognize and act upon new opportunities, and generally improve business performance through efficiency gains. Having quick and specific access to the right information can enable higher quality business decisions that in turn raise revenue and profit. Predictive information governance can provide a foundation on which to finally get Knowledge Management on track.
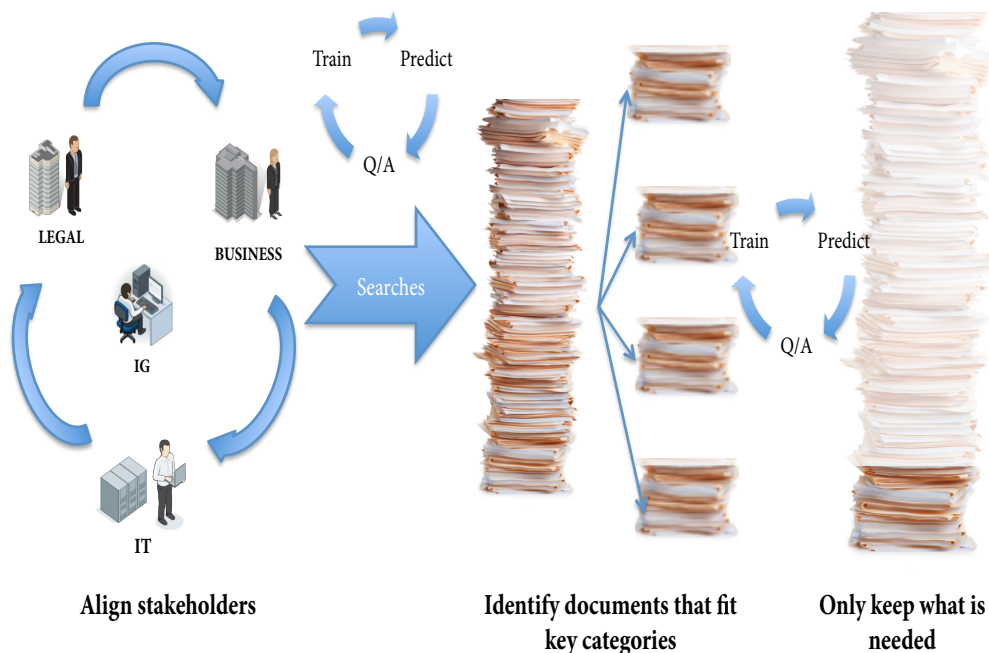
How Does Predictive Information Governance Work?

First, Predictive IG does not eliminate the need for humans. In fact, to optimally leverage the power of predictive information governance, humans are an absolute necessity. Predictive IG marries human intelligence, technology, and statistics in order to predict things about information assets, such as whether they are responsive to a litigation matter, whether they need to be included in a Legal Hold, if they are likely to contain business value, or if they should be disposed of.

Predictive IG can help business users get started in understanding exactly what kinds of information exist. Rather than simply relying on keyword search to filter out documents, predictive technologies combine things like concept clustering, entity extraction, and categorization with human knowledge about business issues. Doing so gets the most relevant documents in front of humans as quickly as possible. Humans can then make decisions about documents (e.g. keep or delete), and the system can iterate on the rest of an information store to predict with other documents should be kept or discarded. What is important to note is that the process requires humans, requires iteration because we keep learning more and more about content as it get reviewed, and requires validation to ensure that the process is working effectively.

The diagram below of the use of predictive information governance in defensible deletion provides a good example.

Using Predictive Information Governance For Defensible Deletion

| Align stakeholders | Identify documents that fit key categories | Only keep what is needed |

Are There Risks In Using Predictive Information Governance?

As with the use of any kind of new technology or methodology, there is the risk of making mistakes that could have negative consequences, such as sanctions for the destruction of information that should have been on Legal Hold. It is important to remember, however, that the standard required by US Federal Rules when it comes to retaining and preserving information for litigation is 'reasonable effort' rather than 'perfection'. As long as predictive IG is used in a way that incorporates human knowledge, has strong quality assurance measurements, and can be mathematically validated, it will likely be seen as a core part of reasonable efforts to conduct IG.

Anecdotally, litigation attorneys that eDJ speaks with regularly point out that the use of predictive information governance in the IG paradigm – for defensible deletion or record classification, for example – is very unlikely to be challenged. While Predictive Coding for Legal Review may be challenged, it is difficult to justify challenging a company's whole retention management program because doing so tends to be outside the scope of a given matter and would blow the matter's budget away. That does not mean that the use of predictive IG would never be challenged, but as long as organizations plan correctly, use predictive technology reasonably, and bring in the right expertise in terms of Q/A, technology, and statistics, the risks should be low.

Leverage Predictive IG To Take Programs To The Next Level

Because effective IG can reduce costs, mitigate risks, and even help grow revenue and create new business opportunities, the time is now to get IG programs running on all cylinders. More often than not, it is smart to tackle IG in discrete projects such as Legal Hold management, defensible deletion, migration from shared drives to managed repositories, or smarter eDiscovery collection and preservation projects. With approximately two-thirds of organizations either having an IG program in place or in the planning stages, now is the time to get it right. Predictive IG can be the bridge from today's immature and lost IG efforts to tomorrow's smart, crisp, effective IG projects that leverage all the two most important tools at our disposal – technology and human expertise – and marry them together in such a way that the whole is much greater than the sum of its parts.

About The eDJ Group

eDJ Group offers unbiased information and pragmatic advice, based on years of experience and proven industry best practices. Whether researching a technology or service solution, conducting an eDiscovery Bootcamp or finding the right expertise to answer your specific questions, eDJ Group is the source for all eDiscovery professionals.

We are committed to helping eDiscovery professionals get the information necessary to excel in their professions, rather than offering legal advice or counsel. We operate with the utmost integrity and commitment to our clients on these guiding principles:

- Independence – All research, reports, advice and services are agnostic and conducted independently without influence by sponsors.
- Highest Ethical standards – All content is honest perspective based on real experience and interactions with thousands of practitioners; detailing both successes and failures without favoritism.
- Pragmatic, Experienced Expertise – All services are conducted by industry experts with decades of experience in eDiscovery and strictly vetted by the eDJ Group founders.

For further information about the eDJ Group and their research, please contact Barry Murphy (barry@edjgroupinc.com) or Jason Velasco (jason@edjgroupinc.com).