Metrics-Based Information Governance

Five Ways to Measure Program Effectiveness

Sponsored by:

Abstract
Measuring the effectiveness and business impact of information governance has always been a difficult task. This whitepaper outlines a metrics-based approach that focuses on results in five key areas – compliance, privacy, disposition, litigation readiness and employee productivity. This method, augmented by technology to classify large volumes of data, can give management real insight and facts to justify funding and ongoing support of information governance programs.

Disclaimer
Contoural provides information regarding business, compliance and litigation trends and issues for educational and planning purposes. However, legal information is not the same as legal advice – the application of law to an individual's or organization's specific circumstances. Contoural and its consultants do not provide legal advice. Organizations should consult with competent legal counsel for professional assurance that our information, and any interpretation of it, is appropriate to each organization's particular situation.
# Table of Contents

**Introduction** .......................................................................................................................... 3

**Measuring Information Governance Effectiveness** ................................................................. 3

**Business Drivers for Information Governance** ....................................................................... 4
  - Enable Regulatory Compliance ............................................................................................ 4
  - Control Private and Sensitive Information .......................................................................... 6
  - Defensibly Delete Unneeded Information ............................................................................ 7
  - Improve Litigation Readiness ............................................................................................... 8
  - Enhance Employee Productivity and Collaboration ............................................................. 9

**The Role of Technology in Metrics-Based Information Governance** ............................... 10
  - Automate the Classification of Information ......................................................................... 10
  - Develop a Metrics Dashboard ............................................................................................... 12

**Conclusion** ............................................................................................................................. 12

**About Recommind** .................................................................................................................. 13

**About Contoural, Inc.** .............................................................................................................. 13
Introduction

Every company has a substantial information governance challenge, involving discovery and legal holds, regulatory compliance and information over-retention. But most organizations – if they execute information governance programs at all – do so largely on faith: millions of dollars can be spent without knowing just how effective their program is.

Organizations need a smarter way to measure the value of information governance initiatives, focused on measuring its benefits in five specific areas. A metrics-based approach answers the question often posed by management: will our information governance effort be effective and will it really help our business?

Measuring Information Governance Effectiveness

Information governance (IG) combines programs for records and information management (RIM), eDiscovery, litigation readiness and the control of private and sensitive information so that information – wherever it resides – is properly created, stored, used, protected and deleted in order to further business goals.

Given the increasing volume of electronically-stored information (ESI), social media and collaboration, trends in mobility and “bring-your-own-device” (BYOD), expectations for greater innovation and employee productivity, increasing litigation and regulation – plus “Big Data” – it’s no wonder that the demands on the people and processes responsible for information governance are amplified.

It is common to find that senior management is hesitant to invest in information governance without having real data to answer some basic questions:

- How much of a problem exists?
- What can be done to address it?
- How can our program effectiveness be measured?
- Does information governance yield material business benefits?

Some organizations have turned to external models to help them establish the value of information governance. These include frameworks for the design and operation of a RIM program, eDiscovery workflow models, and best practices guidelines. The models and frameworks are useful, and organizations certainly should examine them to see how they can help, but we feel that they often lack actionable guidelines to
measure program effectiveness. Instead, what is needed is a metrics-based approach to information governance that is focused on results.

**Business Drivers for Information Governance**

There are several reasons why an organization should adopt information governance.

- **Enabling compliance.** Understand and embrace pertinent legal and regulatory mandates, apply retention policies as appropriate and remove what is not needed.
- **Protecting sensitive information.** Identify what must be protected, such as personally identifiable information (PII), trade secrets and other types of corporate confidential data, and establish guidelines to ensure proper management.
- **Defensible deletion.** Centralize the control of information deletion to defer or avoid IT expenditures, improve application performance and reduce costs in multiple business areas.
- **Optimizing eDiscovery.** Assert control over information before the next legal action and establish repeatable and predictable legal hold processes to minimize business disruption.
- **Improving employee productivity and collaboration.** Help employees to “save smart,” search and locate what they need to improve their job performance.

Many clients see positive results across several of the business drivers in their organizations:

- Defensible deletion has helped improve regulatory compliance and employee productivity.
- Effective, improved compliance capabilities have led to fewer and more easily-managed eDiscovery events.

The measurements presented in this whitepaper reflect our experience across a broad range of enterprises in many industry sectors. Perhaps it’s obvious – be mindful that your results may vary!

**Enable Regulatory Compliance**

In the context of information governance, “compliance” is the requirement that an organization’s policies and processes for retaining, managing, securing and eventually disposing of documents and data adhere to certain legal and regulatory
mandates. Achieving compliance can be challenging for many reasons: applicable regulations depend on the industry, sector (e.g., public or private), and geography in which an organization operates. Rules are subject to changes and updates. And the volume of electronic and paper information that organizations create and use each day makes it costly and time-consuming to identify what should be retained and for how long. Finally, while achieving compliance is required, it is important that doing so does not undermine day-to-day business operations.

Ensuring compliance may require substantial manual effort. Start with a review of the regulatory landscape, assess the organization’s records and information management (RIM) processes, policies, record retention schedules (RRS) and legal hold procedures, and update them as necessary. Next, establish a baseline measurement of the organization’s records. Do this by inventorying and classifying the records that exist: Apply retention policies as appropriate and defensibly delete what is not needed. A metrics-based information governance program can reduce the volume of unmanaged and/or expired records, unneeded information and duplicated data to negligible levels (see Figure 1).

![Figure 1: Effectively Control Records to Enable Compliance](image-url)
**Control Private and Sensitive Information**

Any organization that collects or maintains confidential and sensitive data, including personally identifiable information (PII) and protected health information (PHI), must be sure that such information is secured and remains private. Additionally, organizations must control trade secrets and other types of corporate confidential data to protect against conflict of interest risk and intellectual property leakage. Unfortunately, most privacy activity today focuses on “after the incident” responses to information breaches or leakage.

Our work with clients confirms that private and sensitive information can be found just about everywhere. “Official” repositories that contain confidential data are often clearly identified, centrally managed, secure, and have proper disposition policies in place. But breaches of privacy and confidentiality can arise from many sources, including:

- The “bring your own device” (BYOD) trend and the storage of information on mobile devices.
- Moving information to insecure areas such as file shares and email PST files that lack appropriate access controls.
- Frequent and casual interactions among customers, partners, government agencies and employees.
- The insecure handling and disposition of hard copy, removable media, retired PCs, laptops, systems and servers.

For a 5,000 person enterprise, it is common to find that unsecured confidential information and intellectual property assets comprise as much as 10% of the total amount of stored information. Failure to manage this sensitive data properly can result in penalties, expense and reputational damage.

To protect sensitive information, start by knowing what must be protected. Measure the amount of private and confidential information, which document types are relevant, where they are stored, how they are transmitted and the effectiveness of existing controls with respect to applicable internal policies and external regulations. If needed, develop standards for classification of documents and information in terms of levels of sensitivity or security (such as Public, Internal Use Only, Confidential, etc.). Finally, classify information appropriately, thus enabling data size reduction, optimal placement of data in the storage hierarchy and the proper management of what should be held.
Defensibly Delete Unneeded Information

Continued data growth consistently overwhelms the capability of IT departments to manage it. At a typical 5,000 person enterprise, we’ve found that the amount of expired records, unneeded information and duplicated data exceeds the amount of information with business value and data under legal holds by a factor of three. Such low-value data is found in file shares, backup tapes, Microsoft SharePoint, enterprise databases, and unmanaged repositories such as email PST files and removable media. Such “junk” consumes valuable storage space, reduces the performance of enterprise applications and hampers the ability of employees to search for and find what they truly need.

To overcome the consequences of information overload, organizations must adopt “defensible deletion” policies that allow for systematically disposing of information that is not needed for legal, regulatory and business reasons. These policies demonstrate that reasonable steps are in place to protect the information and the organization itself. Courts have held that such policies must be routine, consistent, transparent, carried out in good faith and applicable across the organization. With defensible deletion policies in place, decisions can be readily understood and explained to non-IT and non-records professionals. A metrics-based approach can reduce the volume of unneeded information and duplicated data to virtually zero. In a typical 5,000-person organization, we believe that as much as 75% of electronic information can be defensibly disposed without incurring the wrath of business units (See Figure 2). An obvious benefit is to help the IT department defer or avoid costs for storage, staff, power and software licenses. As well, important advantages arise in other areas. For example, shrinking the data footprint can reduce the cost of litigation and improve compliance.

<table>
<thead>
<tr>
<th>Putting defensible deletion into place means investing time and effort to</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Measure total data volumes across all media types.</td>
</tr>
<tr>
<td>• Define retention policies that include the business justification and process for holding on to information</td>
</tr>
<tr>
<td>• Create deletion objectives and measurements across all types of media, including email, files, backup tapes, etc.</td>
</tr>
<tr>
<td>• Train employees to “save smart” and monitor ongoing effectiveness.</td>
</tr>
</tbody>
</table>
Figure 2: Defensible Deletion Reduces Storage and Associated Costs

Improve Litigation Readiness

It is practically impossible to forecast the extent and timing of specific, future legal matters. But organizations do have information about their past experience that can help them develop metrics pertinent to likely litigation costs:

- What is the organization's litigation profile, including the frequency of past legal actions and litigation trends in the industry?
- What has been the typical discovery cost of a small case (e.g., employment litigation), a medium matter (e.g., a dispute with a supplier) and a large case (e.g., a class-action lawsuit)? How many of each can be expected in a typical year?
- How much data is easily accessible / somewhat accessible / difficult to access / inaccessible?
- How narrow are typical legal holds?
- How efficient and cost effective are current discovery processes?
- What has been an average review cost per document?

A study by the Rand Institute reported median eDiscovery costs of $1.8 million per case (source: http://www.rand.org/pubs/monographs/MG1208.html). Our experience suggests that for a 5,000 person enterprise, a typical “large” matter involves over 55 TBytes of relevant, discoverable information, some of which is readily accessible while some is not. Associated discovery costs can be in excess of $9.8 million.

Effective discovery cost reduction starts with better data management and disposition practices. Shrinking the data footprint provides easier access to information. It reduces the costs of and time required for collection, review, production and processing significantly (see Figure 3), along with lowering spoliation risk. As well, faster assessment of reduced data and information sets enables early and well-informed decisions about optimal case strategy.
Commit to asserting control over information before the next legal action. Build a map of the data sources that might be subject to legal hold. The map can be used by counsel during steps such as early case assessment, meet and confer, discovery scoping and response planning. Classify information to help determine its value, thus enabling defensible deletion to reduce the data footprint. Finally, develop and put in place repeatable and predictable legal hold processes to minimize future business disruption.

**Enhance Employee Productivity and Collaboration**

Our experience suggests that average employee wastes from four to six hours per week searching for, managing and retrieving relevant and useful information. Multiplied by the total amount of employees who use email, unstructured data and documents across the enterprise, this can represent a significant cost. And because end-users expect information to be readily available on their mobile devices so that they can get more work done and collaborate effectively, the problem is only getting worse.

The absence of effective information governance and employees’ “poor information etiquette” are largely to blame. For example, employees’ desire to hoard all the information they think they need actually works against them as they drown in both useful and useless data. Some emails and electronic documents are printed and saved in hard-copy format, which is easily damaged or destroyed. As well, electronic information is commonly

By centralizing document storage, adopting a data placement strategy and standard naming conventions, and enforcing centralized, defensible deletion, employees can more easily find relevant information. In a 5,000-person company, reducing the time that employees spend searching for information from 6 hours to 2 hours per week returns nearly one million hours of potentially productive time to the workforce.
“silenced” on desktops or personal archives and is effectively lost when an employee is terminated.

One of the best ways to allow employees to store, search and locate what they need is to clear out the clutter of unwanted, unneeded low-value information. First, measure the amount of information that is accessible by employees and the time they spend searching for useful data. Establish programs to encourage a “save smart” (e.g., only what’s needed, in conformance with company policies) mentality to align the behavior of end-users with information governance policies. Finally, enforce centralized control to ensure that more information is accessible to others, less is held in single-employee “silos” and the loss of information due to retirement and employee turnover is minimized.

Using this approach, the time and effort required by employees to search for and manage information can be greatly reduced (see sidebar). Employees can share documents without creating multiple copies, improving collaboration and productivity. And aggregating information into a rich source of corporate memory allows it to be mined and analyzed to provide significant business value today and in the future.

**The Role of Technology in Metrics-Based Information Governance**

Knowing what you have and why you have it are important attributes of a metrics-based information governance program. But building awareness requires planning and effort to inventory and classify data. These activities can be accomplished manually. However, such work is time consuming, tedious, subjective, prone to error and is a frequent cause for complaint among employees who would rather not be bothered. Besides, the sheer volume, variety and disparate locations of information flowing through today’s organizations make most manual efforts impractical.

**Automate the Classification of Information**

Automated tools can be used to sort through and index huge volumes of information that would otherwise drive up the costs and risks of discovery and non-compliance. Autoclassification software requires a description of the categories associated with information to be in place. A good starting point is a records management program with well-defined records classes and retention periods that provide the business criteria upon which categorization and indexing can be conducted.
The accuracy of autoclassification can be measured in two dimensions: completeness (also known as “recall”) and precision. It is complete if the system can correctly and consistently identify documents that should be classified while excluding transitory information. It is precise if the information thus identified is placed in the proper categories.

Autoclassification tools fall roughly into two groups, based on the method employed for “teaching” the system about classifying data:

- Rules-based approaches utilize keywords and synonyms along with Boolean logic that is often associated with search engines to confirm objectively a category match with a content item. The precision and completeness of rules-based systems are good when the information to be classified contains sufficient metadata and/or keywords. But the tools require considerable up-front work and testing in order to reduce the frequency of false positive and false negative matches. Especially for large category schemes, it can be challenging to create new terms or change rules as new significant phrases are identified.

- Predictive coding goes farther than rules-based systems. It is a machine learning approach that uses established statistical models and a set of keyword-rich “exemplar” documents to train the software about the context and meaning of information. With predictive coding, relevant information can be identified for each concept in the category scheme. This is especially useful when there is not enough metadata available or when large collections of information are spread across multiple data sources such as email, SharePoint and fileshares (i.e., content “in the wild”). Predictive coding requires effort to create and update the set of training documents, especially as new content is added. But this can be advantageous, as precision and completeness will improve as the software “understands” more about different types of information. For this reason, predictive coding holds the potential of greater classification accuracy than rules-based approaches.

Both rules-based and predictive coding methods were initially deployed to cull large amounts of information for eDiscovery, but they also help drive decisions about retention, defensible deletion and litigation holds. They can be used to identify and apply policies to private and confidential data and to normalizing the information used by employees in day-to-day operations. Stated differently, automated
classification software covers the range of the drivers for a metrics-based information governance program.

Automated tools have become sufficiently trustworthy to assist humans in their decisions about information categorization or, in some cases, to supplant human intervention altogether. The suitability of a particular technology to a specific information governance challenge depends on the volume of information to be categorized, the desired accuracy of the categorization results, and the amount of effort and expense that an organization is willing to invest. We recommend that organizations take a closer look to see how they can be applied.

**Develop a Metrics Dashboard**

Technology is important not only for managing information, but also for gathering insight into the effectiveness of an information governance program. We recommend that organizations develop a metrics dashboard for each key business driver, looking at the initial state (i.e., before the program is put in place), the desired outcome, and sampling at regular intervals to measure progress. Predictive coding tools can be particularly useful for collecting and displaying such measurements.

**Conclusion**

Policies for retention, placement and deletion, and the technology to enable them, are not the only attributes of sound information governance. We believe that metrics should guide the design and operation of IG programs. This will help organizations

- Prioritize – by focusing on one or more business drivers where the needs are most pressing.
- Define achievable objectives – in the case of litigation readiness, this could be a statement like “…to lower discovery costs by $1 million, we must reduce the amount of relevant documents by 20%.”
- Be proactive – by putting new policies and processes in place before a crisis occurs.

Information governance is important, and it need not be overly-complicated. But some work is required! A metrics-based approach – based on the value derived in one or more of the five key areas – can give skeptical executives the solid data they need to justify investment in an information governance program.
**About Recommind**

Recommind provides the most accurate and automated end-to-end eDiscovery, enterprise search and automatic categorization software available. Based on user input, the software understands the meaning behind what users need, automatically and intelligently predicts the best information available and instantly provides the most relevant results.

The name “Recommind” represents the powerful combination of technology and human intuition. Technology is used to instantly and accurately recommend the correct information or expertise that the user has in mind. Similarly, the Recommind logo signifies the extraction of the exact piece of information or expertise a user seeks from the vast universe of available data.

As a leader in the development of best practices in the areas of eDiscovery, regulatory compliance, records management and data retention, Recommind actively participates in a variety of industry groups focused on ameliorating today’s information risk management issues. In addition to being members of the Big Data Coalition and the EDRM group (Electronic Discovery Reference Model), Recommind co-chairs the Search and XML Schema Working Groups. Recommind is also actively involved in the Sedona Conference, the DESI Workshop, Association of Corporate Counsel, ILTA, Women in eDiscovery, AIIM, ARMA and other industry groups.

To learn more about Recommind, please visit www.recommind.com.

**About Contoural, Inc.**

Contoural is the largest independent provider of Information Governance Consulting Services focused on Records and Information Management (RIM), Litigation and Regulatory Inquiry Readiness and Control of Privacy and other Sensitive Information. We do not sell any products or take referral fees, store any documents or provide any lawsuit-specific “reactive” ediscovery services. In this capacity we are a trusted advisor to our clients providing unbiased advice. We have more than 25% of the Fortune 500 as clients, across all industries, as well as federal agencies and local governments.

We provide a real-world, measurable metrics-based approach that reduces risk, lowers costs, ensures compliance and enables higher employee productivity and collaboration. We believe that creating a consensus across a client’s organization is
a cornerstone to an effective strategy. The company's services encompass all electronically stored information (ESI), including e-mail, as well as paper documents.

With an average of 14 years industry experience, Contoural's team is comprised of attorneys, former compliance officers, records managers, privacy professionals and change management experts who have a deep understanding of the legal, compliance and business requirements for retaining and managing information -- as well as seasoned IT professionals with expertise in document archiving, search, litigation management systems, data classification and data storage, all focused on effective program execution.

Contoural offers services that cover:

- Records Maturity Assessment and Strategic Roadmap Development
- Records and Information Management Policy and Schedule Development
- Email and Unstructured Data Management
- Legal Hold and Discovery
- Privacy and Sensitive Information
- Technology Requirements and Adoption
- Enterprise Behavior Change Management
- Legacy Paper and Data Disposition
- RIM Organizational Development and Governance

With these services, Contoural helps enterprises ensure compliance and reduce risk, while achieving litigation readiness and reducing costs.

Contoural, Inc.
5150 El Camino Real, Suite D30
Los Altos, CA 94022
650-390-0800
www.Contoural.com
info@contoural.com

Copyright © 2013 Contoural, Inc.