How Content Assessment Can Reduce Your Risk and Help Manage Storage More Efficiently

An Osterman Research White Paper

Published February 2010

SPONSORED BY

IBM

OSTERMANRESEARCH

Osterman Research, Inc. • P.O. Box 1058 • Black Diamond, Washington 98010-1058
Tel: +1 253 630 5839 • Fax: +1 253 458 0934 • info@ostermanresearch.com • www.ostermanresearch.com

Executive Summary

Your organization ***generates*** an enormous of business-critical content, including emails, word processing documents, spreadsheets, presentations, customer records stored in CRM systems, production orders stored in ERP systems, PDF files received from customers and various other types of documents. Much of this information is sensitive and/or contains important business records.

You ***store*** this information in hundreds or even thousands of locations, such as email servers, file servers, smartphones, desktop computers, laptop computers and document repositories, to name just a few of the many possible venues.

Add to this the large quantity of information that your organization might have generated in a variety of ***legacy*** systems, plus the orphaned content you might have ***inherited*** from acquired companies.

SOME QUESTIONS YOU NEED TO ADDRESS

Generating and storing such large and growing amounts of data leads to some important questions that decision makers must answer:

- Do you understand everything you have?
- Do you know exactly where it's located?
- If you had to go through a regulatory audit or an e-discovery exercise could you easily access and present all of the information that was required?
- If so, could you produce it rapidly?
- Are you certain that all of your sensitive content is encrypted or otherwise protected from data breaches?
- Are you absolutely sure that you have discarded all of the information that your legal counsel and others recommend you dispose of in order to reduce your corporate risk?
- Are you leveraging as much value from your information as possible?

If the answer to any of these questions is No, your organization is at serious risk on a number of levels: potential violation of data breach statutes, non-compliance with legal best practice, adverse legal judgments, loss of customers, loss of revenue or loss of corporate reputation.

Understanding and, by extension, making decisions on how to better manage your information more effectively can help you to solve a variety of diverse issues, ranging from managing e-discovery more effectively to establishing a master data management plan to performing analytics on your content to gain more value from it.

In the past there has been no effective tool that can help organizations deliver insight on their huge and growing volumes of information to inform intelligent decision making on that content. Today, that is no longer the case – content assessment tools are addressing this need and helping to solve previously unsolved and ignored problems.

In short, the capability now exists to enable you to answer Yes to all of the questions above.

KEY POINTS TO CONSIDER

Any organization should do three things:

- Understand what information it has and where it is stored, including information it holds for third parties.

- Understand the obligations it has for this information in the context of retention, encryption and overall good stewardship of the information assets. A key element here is the decommissioning of information assets that are no longer necessary to preserve and so should not be preserved so as to mitigate risk and cut costs.

- Implement the appropriate policies and technologies necessary to ensure that as much value as possible is derived from information and so that it is managed in compliance with all applicable statutes and legal obligations.

ABOUT THIS WHITE PAPER

This white paper focuses on the need to understand the information assets that organizations possess and the drivers that should motivate organizations to manage their information more effectively. The white paper, which has been sponsored by IBM, also discusses IBM's InfoSphere Content Assessment offering, a capability designed to help an organization explore and understand the content within its information repositories and reveal what assets it possesses, where they are stored, and how to effectively manage them for purposes of compliance and risk mitigation.

The Current State of Information Management

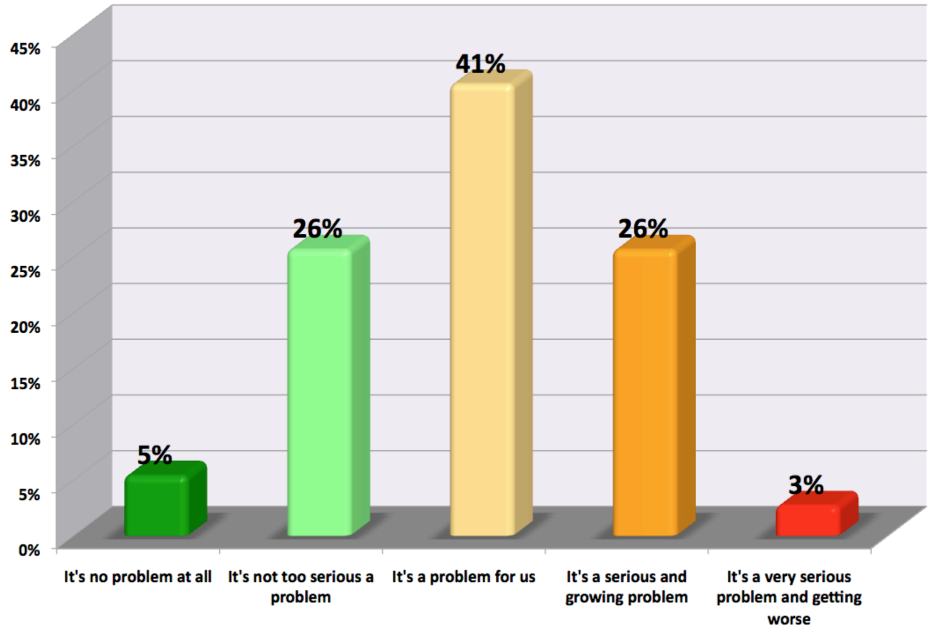
SURVEY FINDINGS

In order to gain a better understanding of the problems faced by organizations in managing their information, Osterman Research conducted a survey in December 2009 to ask organizations specifically about these problems. The results of that research program, conducted with 148 organizations of varying size and across a wide range of industries, is discussed below.

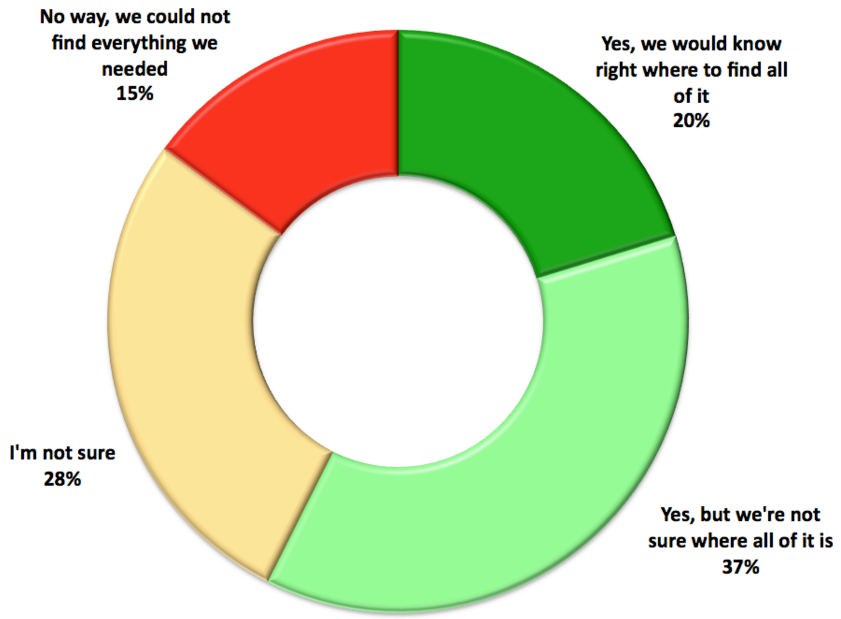
PROBLEMS WITH CURRENT APPROACHES TO INFORMATION MANAGEMENT

As shown in the following two figures, many organizations have fairly significant problems with both the growing quantity of electronic content that they manage, as well as in their ability to be able to find all relevant electronic content when necessary. For example, 29% of organizations find that the growing quantity of electronic content is a serious or very serious problem, and another 67% consider it to be somewhat of a problem. Further, more than two in five organizations either could not find everything they would need for e-discovery or an audit or they are not sure that they could.

Perceived Seriousness of the Problem With Growing Quantities of Electronic Content



Likelihood of Finding All Relevant Electronic Content for Purposes of E-Discovery or an Audit



CRITICAL QUESTIONS THAT ORGANIZATIONS MUST ADDRESS

The data above suggest a number of important questions that organizational decision makers must address in the context of properly managing their information:

- **How much of your content do you have under your control?**
  Most organizations do not have all of their electronic content under control, largely because a) they do not have a good understanding of what they have, and b) they don't know where they could find all of this content if called upon to do so.

  Further, many organizations inherit information over which they previously had no control. For example, if one company acquires another, the acquiring company will likely inherit email servers, other application servers, archives and other content that it did not create. Some or all of this content might be held in ways that do not comply with either the law or with the acquiring company's policies.

- **Do you know your "content topology"?**
  Electronic content can reside in a large number of locations throughout and outside of an organization, including email servers, file servers, collaboration system servers, other application servers, desktop computers, laptops, notebooks, netbooks, CRM databases, archiving systems, etc. The result is that few organizations know where all of their information resides or how they could reliably access it when needed.

  Further, legacy information accumulates throughout an organization, often generated by systems that are no longer in use, are no longer supported by IT or that were deployed by a department or function and never actually under the control and management of IT. Often, this information is locked away in a silo and is not managed in a way that is consistent with corporate policies for security, compliance, risk management, e-discovery, etc.

- **Are you exploiting your information assets as effectively as you can?**
  By understanding what content is available and how it can be accessed, organizations can reuse and leverage their content, thereby gaining more value from these assets.

- **How much does it cost to maintain all of your information?**
  Related to the point above is the fact that few organizations can accurately determine the cost of maintaining all of their information. In short, if decision makers cannot identify where their information is, they cannot accurately calculate the cost of maintaining it.

- **How much space does your content consume?**
  Most organizations cannot determine the amount of space that their information consumes, in large part due to the fact that they don't know how much information they have or where it is located.

- **What are the risks related to your content?**
  Organizations face enormous risk by not knowing what content they have or where it is located. For example emails, files and other electronic content that contains

unencrypted Protected Health Information (PHI) or Personally Identifiable Information (PII) may be held in locations or in a format in which unauthorized parties could gain access to it. This could lead to a variety of problems, including violation of data breach statutes, resulting in fines and other costs. For example, the Ponemon Institute has determined that the cost of a single data breach in FY2009 was $204 per compromised record, up from $138 in FY2005[1]. The Privacy Rights Clearinghouse has chronicled the breach of more than 200 million records since January 2005[2].

- **What content should be managed according to certain policies, but today is not?**
  An inability to understand what information is being held or where it is located leaves an organization vulnerable to managing information in ways that violate regulatory, legal or corporate policies. As part of good corporate governance, it is important to identify the information that poses the greatest risk to an organization – unencrypted and sensitive customer information, confidential communications between senior managers, various types of work in progress, presentations designed only for an internal audience of decision makers, customers' financial information, and protected health information about employees, just to name a few of the many types of content that must be protected. It is critical to find the riskiest types of content and where it is stored and control how this content is managed.

- **Are you storing your information adequately?**
  Organizations are at risk of preserving too much information. For example, if a statute or legal precedent requires information to be held for five years, content held for longer periods exposes an organization to risks during e-discovery or regulatory audits that they otherwise would not have to face. Although not the most serious risk that organizations face, preserving too much information or maintaining it in a more expensive format than necessary drives up the cost of storage. In the survey conducted for this white paper, 21% of organizations reported that storage management is a serious and growing problem. As the quantity of content continues to grow over time, this will become a more serious problem.

- **How much labor can you devote to managing information?**
  Managing information requires investments of labor, particularly for activities like processing e-discovery or when responding to a regulatory audit. By streamlining their information management – including reducing the amount of content that they must store and process – organizations can significantly reduce their overall cost of information management.

Drivers for Improving Visibility Into Your Content

There are a significant number of drivers that should motivate organizations to understand what content they have available and how this information should be

[1] http://news.cnet.com/8301-27080_3-10440220-245.html

[2] http://www.privacyrights.org/ar/ChronDataBreaches.htm

managed. Among the more important factors are those focused on e-discovery and regulatory compliance requirements, although there are many drivers that come into play depending on the industry in which an organization participates, its geographic distribution, etc.

E-DISCOVERY AND LITIGATION SUPPORT

One of the most important drivers for e-discovery has been the set of amendments to the Federal Rules of Civil Procedure (FRCP) that went into effect on December 1, 2006. These changes represented several years of debate at various levels and will have a significant impact on electronic discovery and the management of electronic information within organizations that do business in the United States. The changes to the FRCP require organizations to manage their content in such a way that this information can be produced in a timely and complete manner when necessary, such as during legal discovery proceedings.

The changes reflect the reality that discovery of Electronically Stored Information (ESI) is now a routine, yet critical, aspect of every litigated case. For example, the FRCP amendments treat ESI differently and they require early discussion of and attention to electronic discovery. They also address inadvertent production of privileged or protected materials, encourage a two tiered approach to discovery, and also provide a safe harbor from sanctions by imposing a good faith requirement.

When a hold on information is required, it is imperative that an organization immediately be able to begin preserving all of its relevant electronic content – obviously, this requires an organization to know what information it has and where it can be found.

If an organization is not able to adequately place a hold on information when required, it can encounter a variety of serious consequences, ranging from embarrassment to serious legal sanctions or fines. Litigants that fail to preserve content properly are subject to a wide variety of consequences, including brand damage, additional costs for third-parties to review or search for information, court sanctions, directed verdicts or instructions to a jury that it can view a defendant's failure to produce information as evidence of culpability.

REGULATORY COMPLIANCE

There are a large number of obligations to protect various types of information, among which are:

- **Sarbanes-Oxley**
  The Sarbanes-Oxley Act of 2002 requires all public companies and their auditors to retain such relevant records as audit workpapers, memoranda, correspondence and electronic records – including email – for a period of seven years. Company officers are obliged to report internal controls and procedures for financial reporting and auditors are required to test the internal control structures. Businesses have to ensure employees preserve information – whether paper- or electronic-based – that would be relevant to the company's financial reporting.

- **SEC and FINRA Rules**
  Members of national securities exchanges, brokers and dealers are obliged to preserve all records for a minimum of six years, the first two years in an easily accessible place (SEC Rule 17a-4). The affected records are broad and encompass originals of communications generated and received by individuals within financial institutions, including inter-office memoranda and internal audit working papers.

- **Health Insurance Portability and Accountability Act**
  The Health Insurance Portability and Accountability Act (HIPAA) of 1996 addresses the use and disclosure of an individual's health information. It defines and limits the circumstances in which an individual's protected health information (PHI) may be used or disclosed by covered entities, and states that covered entities must establish and implement policies and procedures to protect PHI. Penalties for violations are up to $25,000 and $1.5 million, depending on when the violations occurred. Further, an individual who knowingly obtains or discloses individually identifiable health information may face a criminal penalty of up to $50,000 and up to one-year imprisonment.

- **Gramm-Leach-Bliley Act**
  The Gramm-Leach-Bliley Act requires that financial institutions protect information collected about individuals, including names, addresses, and phone numbers; bank and credit card account numbers; income and credit histories; and Social Security numbers.

- **Payment Card Industry Data Security Standard**
  The Payment Card Industry Data Security Standard (PCI DSS) encompasses a set of requirements for protecting the security of consumers' and others' payment account information.

- **Regulation S-P**
  Regulation S-P has been adopted by the US Securities and Exchange Commission (SEC) in accordance with Section 504 of the GLBA. This section requires the SEC and a variety of other US federal agencies to implement safeguards to protect non-public consumer information, and to define standards for financial services firms to follow in this regard.

- **Personal Information Protection and Electronic Documents Act**
  The Personal Information Protection and Electronic Documents Act (PIPEDA) is a Canadian privacy law that applies to all companies operating in Canada. Like many other privacy laws, it requires that personal information be stored and transmitted securely.

- **Family Educational Rights and Privacy Act of 1974**
  The Family Educational Rights and Privacy Act of 1974, which is focused on protecting the privacy of students' education records, includes provisions for how states can transmit information to Federal entities.

- **UK Data Protection Act**
  The UK Data Protection Act imposes requirements on businesses operating in the United Kingdom to protect the security of personal information and to preserve information only as long as it necessary to do so. The Act requires, at least by implication, requirements for encrypted transmission of personal information and its secure retention.

- **Model Requirements for the Management of Electronic Records (MoReq)**
  MoReq, originally developed in 2001, defines the functional requirements for the manner in which electronic records are managed in an Electronic Records Management System. MoReq has been used widely in Europe and has been updated with MoReq2.

All of these regulations can impact an organization's information lifecycle governance requirements to varying degrees, dependent largely on the industry it serves. For example, broker-dealers in the United States will need to focus heavily on SEC and FINRA rules, educational institutions will need to focus on FERPA requirements, and organizations in the consumer financial services space will need to focus on PCI DSS rules. Various regulations will have an impact on how information is protected and managed.

RISK MANAGEMENT

One of the key benefits of managing information more effectively is to help an organization to manage its risk more effectively. Because holding information that may be sensitive or that may be sought by others carries with it inherent risk, understanding where content is located, what policies need to be applied to it, and how it is to be managed can significantly reduce the risk that an organization faces.

As a result, organizations can manage their risk more effectively by more fully understanding the information assets they have in their possession. By understanding what content has value to the business, what content does not have value, and how to retain the important content and dispose of the rest, organizations can apply effective lifecycle governance to their content.

CONTENT DECOMMISSIONING

Content decommissioning – disposing of content that is no longer necessary to maintain – is a critical best practice for any organization for a variety of reasons:

- Content archiving can be more efficient simply through the reduction of bytes that must be indexed and archived.

- Minimizing the amount of content available for pre-litigation review, an e-discovery exercise or a regulatory compliance audit will reduce the amount of content that needs to be searched, processed and analyzed. This can significantly reduce the legal and regulatory compliance costs associated with these efforts and could potentially save an organization hundreds of thousands or millions of dollars annually.

- Storage costs can be significantly reduced simply by reducing the number of terabytes that need to be managed. Cost savings are derived from things like lower labor costs from a reduced number of storage systems that must be managed, reduced power costs and reduced cooling costs.

- Reducing the amount of content on email servers, collaboration servers, file servers and the like can improve the performance of servers, shorten backup windows and reduce the amount of time required to restore a server in the event of a crash.

Next Steps

There are three things that any organization should undertake as it attempts to minimize its risk and improve its overall information management strategy:

- **Business drivers**
  First and foremost, you must understand why you need to understand your content. In other words, what are the drivers for managing information assets more effectively: the long-term health of the business, minimizing exposure from regulatory audits, reducing legal exposure, generating more revenue, creating more value from your information, or all of the above? Related to this is the need to prioritize the drivers and make decisions accordingly.

- **Quantify the value of decommissioning your content**
  Insofar as possible, quantify the value that you might realize by decommissioning the content that you no longer need. This might involve a return-on-investment analysis or some other calculation of the business value that your organization can realize by eliminating unnecessary content.

- **Assessment**
  Your organization needs to understand the information in its possession (including content that is held for other parties). An organization of any size will need a system to explore and analyze the content in its information repositories and provide an assessment of the information it has and where it is located.

- **Compliance obligations**
  You will need to know the regulatory, legal and other obligations your organization must satisfy and how these requirements might change in the future. This will depend to a large extent on the advice of your internal and external legal counsel; continuing review of recent court rulings and regulatory agency orders and rulings; the states, provinces and countries in which your organization operates; the industry your organization serves; etc. All of this information, in turn, will be used to establish corporate policies as part of an overall governance framework for managing information more effectively and minimizing risk.

- **Executing on governance requirements**
  Once you know what content you have and your obligations for protecting that information, you will need to manage it effectively: encrypt some content types in

transit and at rest, archive some content for the appropriate length of time, delete information so as to mitigate your corporate risk, etc. These are critical elements of a lifecycle governance system for current information assets, as well as assets that the organizations will develop and procure in the future.

The bottom line is that organizations must first understand the data they possess and where it is located. Then, based on the variety of business drivers that impact their organization – legal, regulatory or simply best practice – they must govern their data appropriately.

About IBM's InfoSphere Content Assessment

Practically every organization is struggling with growing volumes, variety and velocity of information. IBM is offering a solution that can help organizations solve this problem through dynamic analysis of their content. Leveraging over a decade of content analytics research and technology, IBM offers InfoSphere Content Assessment, which helps organizations to tame the ever-growing content sprawl, to save costs, reduce compliance risks, and enable their content for better business leverage.

Using content analytics, this new offering helps organizations rapidly assess unmanaged content, bloated production systems, or legacy content-based systems to detect unnecessary, over-retained, irrelevant, or duplicate content that is eligible for decommissioning. This creates the opportunity to dramatically reduce production system burdens and associated storage requirements. By deriving insight about their otherwise uncontrolled content, this offering can also help identify high-risk and valued content like corporate records that need to be secured, retained and managed throughout its lifespan.

InfoSphere Content Assessment focuses on three key areas:

- **It dynamically analyzes what you have**
  Aggregate, correlate, visualize and explore your enterprise information in new ways to understand virtually all content types from multiple sources. Make rapid decisions about the business value, relevance and disposition of the information located throughout your organization.

- **It decommissions unnecessary content**
  Save costs and reduce risk by eliminating obsolete, over-retained, duplicate and irrelevant content – and the infrastructure that supports it.

- **It helps your organization preserve and exploit the content that matters**
  Collect valued content to manage, trust and govern throughout its lifespan in an enterprise-grade electronic content management platform. Uncover new business value and insight by integrating with solutions for e-discovery, information lifecycle governance, case management, and more.

IBM is the first to market a content assessment tool that can explore multiple content sources and use analytics to deliver new business understanding and visibility from the content and context of unstructured information. This allows organizations to take appropriate actions on its content, such as decommissioning of unnecessary content and legacy content sources, or dynamically collecting information to perform ad-hoc e-discovery collection requests.

Summary: A New Way of Looking at Content

What has been described in this document is a new way of looking at content management. Instead of simply viewing information management as a necessary part of the overhead cost in an organization, what we're describing here is an approach that does several things:

- It helps an organization to understand the information assets it possesses and where this information is located in a way that traditional information management tools cannot do.

- By knowing what information is available and how it can be accessed, organizations can leverage this content to make faster and better-informed business decisions.

- Content assessment can help an organization to reduce its overall stockpile of information, reducing its storage costs and making access to necessary information more efficient and faster.

- Risk can be mitigated by eliminating content, managing sensitive information in compliance with corporate policies, regulatory statutes and legal obligations.

In short, organizations that implement appropriate content assessment practices and technologies will realize a significant return-on-investment from the additional value they will derive from their information assets and from the reduced risk they will incur.

Osterman Research, Inc. does not provide legal advice. Nothing in this document constitutes legal advice, nor shall this document or any software product or other offering referenced herein serve as a substitute for the reader's compliance with any laws (including but not limited to any act, statue, regulation, rule, directive, administrative order, executive order, etc. (collectively, "Laws")) referenced in this document. If necessary, the reader should consult with competent legal counsel regarding any Laws referenced herein. Osterman Research, Inc. makes no representation or warranty regarding the completeness or accuracy of the information contained in this document.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND. ALL EXPRESS OR IMPLIED REPRESENTATIONS, CONDITIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE DETERMINED TO BE ILLEGAL.