WHITE PAPER

Developing an Information Governance Strategy

An Osterman Research White Paper

Published February 2014

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

EXECUTIVE SUMMARY

Electronic information – emails, files, word processing documents, spreadsheets, instant messages, presentations, social media posts, text messages, databases, etc. – are burying organizations in a rapidly expanding information avalanche. Most of the data piling up in organizations is of the unstructured variety also known as "data in the wild" as opposed to structured data. The vast majority of this unstructured data is ignored by the organization, left for employees to manage (or not), and is almost never accessed ever again. This unmanaged and abandoned data will probably never see the light of day again – if it does, it might be in ways that are damaging to an organization, such as during litigation or as a surprise in a regulatory audit.

UNCONTROLLED, UNMANAGED INFORMATION

This uncontrolled content growth is making it almost impossible for organizations to adequately capture, manage, store, share and dispose of information in any meaningful way. As organizations put off addressing this growing data problem, they are quickly reaching a point of no return. This unmanaged information build-up raises the cost of storage, negatively impacts employee productivity, and raises the cost and risk of eDiscovery.

Successful organizations run on, and are dependent on, the creation and consumption of information. But information is valuable to an organization only if decision makers and others that need it know where it is, what's in it, what is shareable and by whom it is shareable…in other words, the need is for *managed* information.

In the past, organizations have relied on end users to decide what should be kept, what should be deleted, where it should be stored, and for how long it should be retained. In fact 75% of data today is generated and controlled by individual employees. In most cases, this practice is ineffective and causes what many refer to as "covert" or "underground archiving", where individuals end up keeping everything in their own unmanaged local storage repositories or archives. These underground archives effectively lock most of the organization's information away, hidden from everyone else in the organization and give birth to the phenomenon called "dark data" – an underground of unmanaged and uncontrolled unstructured data. Inexpensive, cloud-based storage has only served to exacerbate the problem of dark data.

Information is valuable to an organization only if decision makers and others that need it know where it is, what's in it, and what's shareable.

Nowhere is this problem more evident than in the widespread and growing use of file sync and share tools, most notably Dropbox. While these tools are useful for individual users, they allow corporate policies, legal requirements and regulatory obligations to be circumvented, creating a serious information governance problem. Organizations that solve even just this one issue have taken a significant step toward solving their overall information governance problem.

The lack of good information governance has brought us to an inflection point: decision makers must gain control of their information to enable innovation, profit and growth; or continue down the current path of information anarchy and potentially lose out to competitors who are better able to govern their information.

ABOUT THIS WHITE PAPER

This white paper discusses the variety of challenges focused on information governance and also offers a variety of recommendations about what organizations can do to improve their information governance practices. The paper also provides a brief overview of its sponsors – Actiance, ownCloud, Rand Secure Data and StorageMadeEasy – and their relevant solutions.

WHAT IS "INFORMATION GOVERNANCE"?

There are many definitions for Information Governance in use today, some better than others. The one most referenced is:

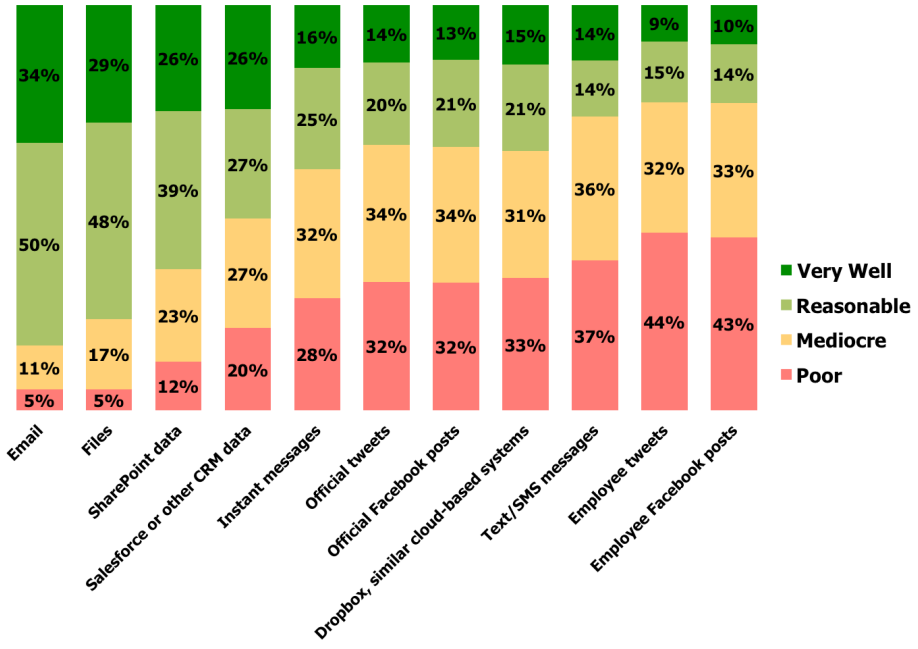
> *"Information Governance is a cross-departmental framework consisting of the policies, procedures and technologies designed to optimize the value of information while simultaneously managing the risks and controlling the associated costs, which requires the coordination of eDiscovery, records management and privacy/security disciplines."*

This definition encompasses the important points of an effective information governance program. The key section of this definition of information governance is the statement that it is made up of *policies, procedures and technologies.* Technology is an important piece of an effective information governance program, but without well thought-out, documented policies and tested procedures, an information governance program can't be successful. Let's look at the three definition focus areas in more detail.

THE CURRENT STATE OF INFORMATION GOVERNANCE

To determine the current state of information governance, Osterman Research conducted a survey with organizations across a wide range of industries. Our goal was to determine how well decision makers and influencers believe their organization is managing their governance of various types of content. As shown in Figure 1, most organizations believe that they do a good job when managing email, files and SharePoint data. For other data types, however, a growing proportion of organizations believe that their governance practices are mediocre or poor. This clearly indicates that information governance, on balance, is sorely lacking, particularly for less traditional data types like social media posts or text messages.

Figure 1
Quality of Information Governance



The information governance policy should explain why the policy was created and why it's important to the organization.

INFORMATION GOVERNANCE POLICY

An information governance policy provides guidance on how organizational information value can be maximized by efficient and ethical handling of its content by ensuring that information is:

- Held securely and confidentially
- Obtained fairly and lawfully
- Stored securely, accurately and reliably
- Used effectively and ethically
- Shared and disclosed appropriately and lawfully
- Audited with regard to who has access to it
- Disposed of per regulatory instructions

The information governance policy should explain why the policy was created and why it's important to the organization, as well as an explanation of how information is used. The policy should include a designated person to direct questions to, as well as their contact information. The information governance policy should also list what are unacceptable information use/management and the penalties that will be administered for not following the policy.

Laws and regulations change on a regular basis. Because of this, the information governance policy should be periodically reviewed and updated to ensure it continues to meet all legal and regulatory requirements and best practices.

INFORMATION GOVERNANCE PROCEDURES

The actual implementation of the organization's information governance policy (processes and procedures) is where approved, documented and tested actions actually cause the information to be properly governed. Examples of prescriptive procedures include:

- All information subject to regulatory retention should be moved/copied to the department share drive under the "compliance" folder within 36 hours after finishing with it.

- All final contract agreements should be moved to the department SharePoint repository (or equivalent) within seven days after receipt and filed in the "contract" folder and tagged with the appropriate retention period.

- Any documents that contain personally identifiable information (PII), should be marked as such and secured in the appropriate repository as soon as possible.

Information governance procedures tell the individual (or technology solution) what should happen to information and how it should be handled.

An information governance audit process is an important part of the process. It helps quickly answers questions about its efficacy and can suggest changes for fine-tuning, such as:

- Are employees following the policy?

- Are they developing work-arounds?

- If so, why are they developing work-arounds? Is it because the procedure is needlessly complex, etc.?

- Can additional automation relieve the problem?

An on-going audit process is the best way to determine the efficiency and effectiveness of the policy and procedures. It also adds to the defensibility of an

Information governance procedures tell the individual (or technology solution) what should happen to information and how it should be handled.

organization's policy and procedures in the event a problem arises during litigation or regulatory inquiry.

INFORMATION GOVERNANCE TECHNOLOGY

In today's digital world, businesses are dependent on timely access to the right information. And because today's modern business has to deal with huge amounts of information on a daily basis, businesses are being forced to abandon the manual process of relying on employees to actually manage all of their information. Today, employees simply don't have the time or ability to interpret and follow retention schedules that are sometimes hundreds of pages long. By necessity, they are being pushed into adopting computer automation and software to get ahead of, or at least stay current with, the information flowing throughout their infrastructures on a daily basis. There will be a more detailed discussion about information governance technology later in this paper. Also, later in this white paper we will discuss some of the key information governance technologies.

RECORDS MANAGEMENT VS. INFORMATION GOVERNANCE: WHAT'S THE DIFFERENCE?

THERE ARE OVERLAPS, BUT THEY ARE DIFFERENT

Records Management (RM), also known as Records Information Management (RIM), is the practice of managing (controlling and governing) documents, files or other content that are designated by a regulatory authority (federal or state government and sometimes legal authorities), to be proof of specific business activities and therefore "records of the business". Records management includes identifying, classifying, prioritizing, storing, securing, archiving, preserving, retrieving, tracking and finally disposing of company records when allowed to by law. Until several years ago, the vast majority of company records were in hardcopy form, making it relatively easy to capture and manage these records. In today's business environment, more than 95% of all organizational information is created, consumed and disposed of in a digital format. This may seem like an easier format to control and manage, but in reality it greatly complicates the management of information because digital information can be created, consumed, moved, edited and deleted in a matter of seconds.

In the past, an organization could get away with ignoring the vast majority of electronic data as transitory and concentrating on only hardcopy records. For those small proportion of electronic documents there were considered a record, organizations instructed individuals to "print them out" and file them as they would hardcopy records – destroying any metadata attached to them. Now with almost 90% of information workers spending their day on their workstations, laptops, and smartphones creating work documents and presentations in office productivity software, communicating with email and instant messages, storing content in consumer-grade file sync and sharing tools, and conducting marketing activities on social media, "records management" is not as straightforward and hardcopy-based as it once was.

Add to the above argument the appearance of "big data analytics"; the process of examining huge amounts of data (terabytes to petabytes) to uncover beneficial hidden patterns, unknown correlations and other useful information. With so much activity around digital information, just managing "business records" for regulatory requirements becomes almost secondary.

Unlike current records management solutions, information governance is a superset that *includes* records management. An information governance solution ideally provides for the enterprise-wide identification/indexing of all information, the centralized management of all information via retention/disposition policies, and automated information sharing based on security levels while supporting the

Unlike current records management solutions, information governance is a superset that includes records management.

enforcement of information governance policies across business functions, locations, and information silos.

Today's truth is that Records Management is now a relatively small, but still important subset of the bigger Information Governance need.

ADDITIONAL INFORMATION GOVERNANCE SUBSETS

Records Management

Records management activities are normally centered on regulatory data retention laws that require organizations to keep and manage specific documents (or records of the business or activities) for required amounts of time under penalty of fine, loss of business, loss of regulatory license, or in rare cases, jail time. Specifically, records management consists of building and managing to a framework by which organizations manage the collection, retention, and eventual destruction of specific "records of the business" based on prescriptive requirements from the regulating agencies. The practice of records management almost always targets these compliance records and is not concerned with organizational non-records.

Archiving

Various stakeholders often perceive information archiving differently. To the CIO and IT, archiving is a storage-tiering strategy (HSM) to better manage the cost of electronically stored information (ESI). Rarely opened and less valuable information is stored on low cost media, such as tier 4/5 spinning disc or even tape for longer periods of time. The Chief Compliance officer requires archiving that can be controlled, located, indexed and managed according to retention mandates, both corporate and regulatory. The General Counsel requires that if information is being held, it is held in locations that they can quickly search and from which they can retrieve potentially responsive content. Many general counsels have become conservative and insist information be held for long periods of time so that there is no question about inadvertent deletion and spoliation in litigation. In these cases, ultra-low cost archives are employed to store information for years or tens of years.

Content Management

Content management is defined as the administration of digital content throughout its lifecycle, from creation to permanent storage or deletion. The content involved may be images, video, audio and multimedia, as well as text. Content management repositories are normally centrally located and administered with varying degrees of security and access rights. On the other hand, an information governance process goes further by also attempting to optimize the value of information to an organization through active management and sharing.

eDiscovery

The eDiscovery process involves the identification, collection, review and production of content that could be/is relevant to a given civil lawsuit. The process of identification and collection in the eDiscovery process is highly dependent on the organization being able to quickly find all potentially relevant content. Depending on the judge assigned to the case, the collection process can be extremely time constrained, meaning an organization might have only 30-60 days to actually find, secure and review relevant content no matter where it may exist. These days many organizations responding to an eDiscovery request can find themselves collecting terabytes of data that must be sorted through and many times can miss responsive content altogether – raising the risk of a fine or penalty, or worse, an adverse inference instruction.

Over the last several years, there has been a convergence of eDiscovery responsibilities and information governance activities. For litigation, an effective information governance capability greatly reduces the risk of insufficient eDiscovery response by ensuring all relevant content can be found, secured and reviewed quickly.

Over the last several years, there has been a convergence of eDiscovery responsibilities and information governance activities.

File Sync and Sharing
The pervasiveness of Dropbox and similar tools used to manage corporate information underscores the importance of improving file sync and sharing practices as a key element of good information governance. Bringing this content back under IT control by replacing consumer-grade file sync and sharing with IT-managed alternatives – or, alternatively, enabling Dropbox and other tools to be securely controlled and audited – is a significant step toward improving overall information management.

Information Collaboration
Collaboration, or the working together on a common task or series of tasks, is often dependent on the availability of high quality information across the group. Collaborative tools used in today's businesses are systems that include as one of their major goals the capability to facilitate work that involves more than one person. Examples of these collaboration tools include:

- Email and instant messaging.
- Content-sharing tools, such as SharePoint or departmental share drives. Included here are file sync and sharing tools like Dropbox, Google Drive, Microsoft OneDrive (formerly SkyDrive) and other solutions that are designed to facilitate not only personal access to content, but also sharing of this content with others.
- Other tools, including conferencing solutions like conference calling, WebEx or GoToMeeting; as well as group interactive or social networking applications like Facebook, Google+, and LinkedIn.

All of these tools are designed, to some degree, to facilitate file sharing across multiple devices, logging and auditing of files among users, versioning, collaborative editing, etc. However, while enterprise-grade tools facilitate file-sharing and simultaneously support information governance best practices, consumer-grade tools in which content is managed by employees outside of IT control can be support sharing while actually making information governance worse.

Effective collaboration depends on access by the workgroup on the right information at the right time. If the workgroup is unable to find the information needed to proceed, the collaborative effort is based on old, out of date, or incorrect information. The group work product may then be valueless and a waste of everyone's time. A well-run information governance program ensures the right information is kept and made available for effective collaboration.

Information Sharing
Information can be a valuable asset if it's managed appropriately. The sharing of information can increase the value of that information greatly. Effective information sharing is dependent on a few key points. They are:

- Information should be managed so that like content is linked.
- Information should be indexed so that it can be easily found when searched.
- Information should be classified by access authority so that only employees that meet specified access levels are able to retrieve it.
- Based on the same access controls, information should be easily shareable among others so that work already completed is not duplicated.

Information should be classified by access authority so that only employees that meet specified access levels are able to retrieve it.

KEY INFORMATION GOVERNANCE DRIVERS

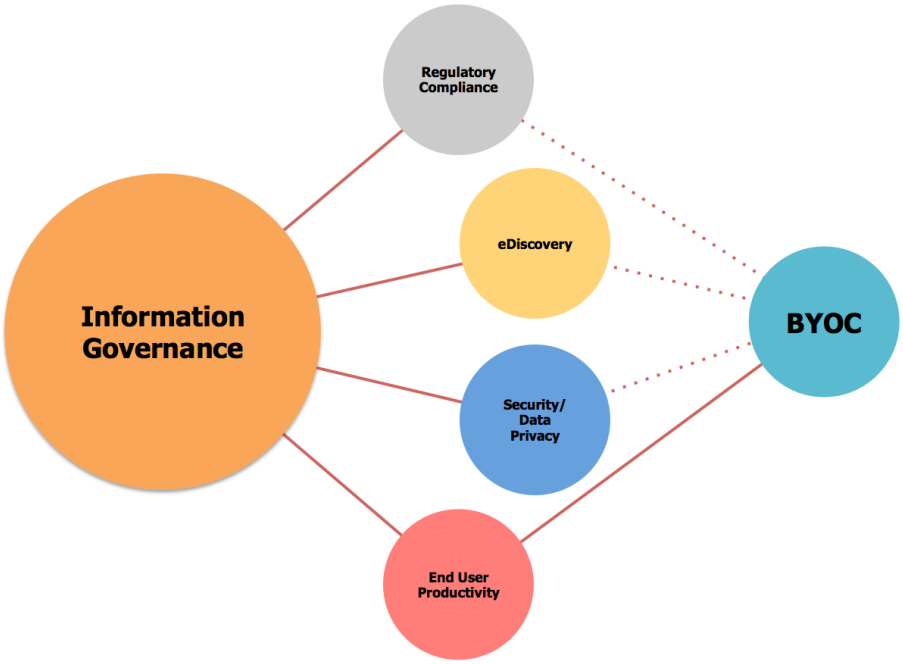
Organizations look to information governance programs to solve a variety of problems normally revolving around risk, cost or both. In many cases, these problems aren't

obvious until they're already negatively impacting the organization. The key business drivers organizations are reacting to are:

- Regulatory compliance requirements - risk
- eDiscovery requirements – risk and cost
- Information Security/Data Privacy – risk
- End-user productivity – cost
- The emergence of public cloud storage (BYOC)

The relationship between these elements is portrayed in Figure 2.

Figure 2
Key Information Governance Drivers



An effective information governance capability greatly reduces the risk of insufficient eDiscovery response.

How can an information governance program address the five key drivers listed above?

REGULATORY COMPLIANCE

Organizations are dealing with increasing numbers of international, federal, state and local regulations that are driving the need to better manage information. Examples of these regulations include SEC 17, FINRA 3010/3011, Solvency II, Dodd-Frank, FAR, HIPAA, the Gramm-Leach-Bliley Act, Sarbanes-Oxley, Bill 198, Basel III, employment-based regulations, and upcoming changes to the European Union Data Protection Directive.

In the 2013 *AIIM Industry Watch: Information Governance* report[i], 49% of survey respondents indicated that they had experienced issues with regulatory authorities due to a lack of complete electronic information. Because of the increase in enforcement actions, general counsels have moved towards retaining everything – forever. The main problem with this strategy is that keeping information is only half of the regulatory requirement. An organization still must find the information when it's requested. If this content is spread across the enterprise, much of it controlled by employees in cloud-based data repositories and other venues, how do decision makers find it quickly? An effective information governance program will insure that

regulatory information is retained for the proper time period, in the proper storage locations, and retrievable quickly.

eDISCOVERY

For many, the eDiscovery process has become extremely costly and painful. The reasons are not always obvious to those yet to be "discovered", primarily because they don't understand the legal responsibilities and implications of not conducting a rigorous eDiscovery process. Under the existing Federal Rules of Civil Procedure (FRCP) an organization has a duty to begin preserving any and all *potentially* relevant information that could pertain to a case when litigation can be *reasonably anticipated*. The consequence of this responsibility means that decision makers must anticipate future litigation, guess at what information could be relevant to it, and then start gathering and securing for it litigation hold immediately. If not done in an expedient and good faith manner, huge fines, potential loss of case, and responsibility for all court costs – not to mention adverse publicity – are possible.

The other fact that many aren't directly aware of, even those who have been through it before, is the actual cost of eDiscovery. According to the RAND Institute for Civil Justice in their report[ii] entitled *Where the Money Goes*, the document review process consumes more than 70% of every eDiscovery dollar. This fact is important when discussing information governance (or the lack of it) because organizations create/receive and stockpile so much digital data, the amount of reviewable content for even a single lawsuit can easily reach into the millions or billions of pages of reviewable content. A respected U.S. Magistrate, Judge Andrew J. Peck, stated in a video interview on February 4, 2013:

> *"Part of the reason eDiscovery is so expensive is because companies have so much data that serves no business need. Companies are going to realize that it's important to get their information governance under control to get rid of all the data that has no business need… in ways that will improve the company's bottom line…"*

As was stated earlier, there is an obvious convergence of eDiscovery and information governance requirements. An effective information governance capability greatly reduces the risk of insufficient eDiscovery response by ensuring all relevant content can be found and reviewed quickly.

SECURITY/DATA PRIVACY

Employees now have access to gigabytes or terabytes of organizational information with just the click of a mouse button. Ensuring sensitive information doesn't fall into the wrong hands has become extremely challenging for all sizes of organizations as the volume of information continues to grow out of control.

Information security is comprised of two main targets: intellectual property (IP) and trade secrets and customer/employee personally identifiable information (PII):

- IP and trade secrets leakage is a huge risk for companies because this information is easily copied and moved, and many companies don't have any data loss prevention (DLP) solution to address it. Huge investments in IP and organizational know-how can be lost in the blink of an eye or the click of a mouse.

- Protecting PII can be tricky, as well. Many industries have strict federal and state compliance rules for securing and controlling access to PII. Categorization and Entity Extraction technology can be used to securely identify and redact sensitive information within content. New technologies on the horizon include the automatic anonymization of PII for those without the proper access levels. Examples of PII include Social Security numbers, bank and brokerage account numbers, personal addresses, driver's license numbers, and Personal Health Information (PHI).

Information security is comprised of two main targets: intellectual property and trade secrets and customer/employee personally identifiable information.

Data leakage is a critical concern for most organizations as sensitive corporate information is migrated to employee-managed clouds and so leaves the control of corporate IT. Sensitive information (usually digital) can be stolen with the click of the "Send", "Tweet" or "Share" buttons as content is copied to the public cloud or elsewhere outside of IT's control. Information governance programs can ensure this data is automatically secured and available only to those with approval to utilize it.

END USER PRODUCTIVITY

Information governance affects user productivity in two ways. First, employees spend time "managing" their work files, their contacts, and especially their email and attachments. This management time includes reviewing content, deciding on whether a file or email should be kept or deleted, and the movement to the final storage location. Many research organizations and experts have advised that this management time is estimated at anywhere from two hours to four hours per week. This may not sound like a great deal of time, but assuming a conservative example of two hours per week for this activity translates to 104 hours per year per employee, or for an organization of 7,000 employees, 728,000 hours per year used for "managing data".

A fully automated information governance solution will reduce this employee management time to near zero because of the assumption that this data would be automatically captured, indexed, managed, stored and disposed of relieving employees of this obligation. Moreover, this underscores the importance of making the right technology choices as a means to regain control over corporate information. Decision makers must consider all aspects of information governance technology – archiving, content management, eDiscovery, file sync and sharing, etc. – in the context of how to improve end user productivity.

The other measure of lost employee productivity is in the number of hours per week employees spend searching for information within the enterprise. Most organizations don't actively manage employee workstations, file shares or even email boxes, and so federated searchable indexes are not available to employees. Employees fall back on going to each potential storage location and performing a simple keyword search, which rarely produces the content the employee is looking for quickly or at all. In many cases the information is not found because of incorrect search terms, questionable file naming practices, or the fact that the content wasn't actually saved at all. In a small percentage of cases, the data can't be found, even though it does exist, and so must be recreated.

If content is not available from every device that employees use to do their work this can contribute to wasted time and effort.

Also important to note is that if content is not available from every device that employees use to do their work this can contribute to wasted time and effort. For example, an Osterman Research survey[iii] conducted in January 2014 found that the typical information worker employs an average of 2.5 mobile devices in addition to their desktop computer, home computer and any other devices they might use to access content. All content that employees might require must be available from all of these devices in order to maximize employee productivity.

With an information governance program and related technology in place, these wasted hours managing, searching for and recreating lost content can be drastically reduced.

BRING YOUR OWN CLOUD (BYOC)

Organizations are now dealing with a new problem, one with potentially enormous liabilities: "free to the public" cloud-based storage repositories. Bring Your Own Cloud (BYOC) refers to the availability and use by employees while at work of free, cloud-based storage space available from companies like Microsoft, Google, Apple, Dropbox and Box, as well as a large and growing number of other vendors. These services provide relatively large amounts of cloud storage space (two gigabytes to 20 gigabytes) at no charge while offering a lightweight desktop and/or mobile client to enhance usability and access to content stored in the cloud.

The advantage to users of these services is the ability to move and store business-related work files that are immediately available from anywhere, including from home, while they're traveling, or even from non-business owned computers. This means employees no longer have to copy files to a USB stick, or worse, email work files as an attachment to their personal email account. The risk of employees using these services is that corporate information can easily migrate away from the organization with no indication that files were ever copied or moved. This also means that potentially responsive information is not protected from deletion, nor is it reliably available for review during litigation.

Stopping employee access to outside public clouds is a tough goal and may negatively affect employee productivity unless the organization offers the same usability that the organization can manage and access as well. For example, several companies have begun offering to employees corporate approved and owned Dropbox accounts with the understanding that the company has direct access to them for compliance, eDiscovery or security reasons, all the while providing the employee the advantages of a cloud account. This also means that the organization can implement information governance policies in the employee cloud. Many organizations are also implementing more robust, business-grade file sync and share capabilities that are offered by a growing number of vendors.

Another approach to "the Dropbox problem" is to implement an on-premises capability that will enable integration with cloud-based services. This approach solves three key problems in the context of BYOC:

- It provides the same ease of use for end users who employ Dropbox and similar types of tools so that users have access to all of their content from any location or any device.
- It puts corporate data behind the firewall so that it can be managed properly by IT in accordance with information governance policies.
- It eliminates the problem of losing data if a cloud provider ceases operation. While these situations or not common, the recent Nirvanix shutdown[iv] and the recent discontinuation of Symantec's Backup Exec Cloud caused angst – and data loss (among Nirvanix customers) – for many.
- It can provide automated backup and audit of any cloud accounts used.

There are many negative consequences from poor or no information governance.

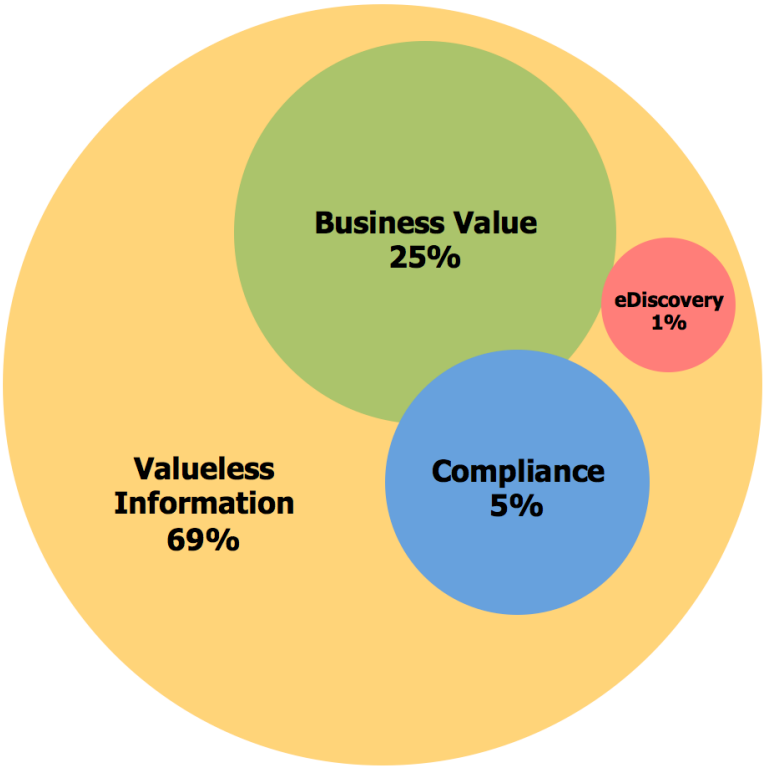
THE RISKS OF POOR INFORMATION GOVERNANCE

There are many negative consequences from poor or no information governance. The number one risk most point to is out of control litigation costs (eDiscovery) and associated fines/penalties from insufficient eDiscovery response. Another risk is the loss of customer confidence or bad publicity from data loss (see the 2013 Target customer data loss case[v]). Other risks include loss of IP or other confidential information, followed by inability to respond to information requests and regulatory action from loss or exposure of PII. The main point to remember is that unlike only a couple of years ago, the consequences of poor information governance far outweigh the cost of effective information governance.

CONTENT DELETION – DEFENSIBLE DISPOSAL

The Compliance, Governance and Oversight Counsel (CGOC) conducted a survey in 2012 that showed that, on the average, 1% of organizational data is subject to litigation hold, 5% is subject to regulatory retention requirements and 25% had some business value, as shown in Figure 3. This means that approximately 69% of any organization's retained data has no business value and could possibly be disposed of without legal, regulatory or business consequences.

Figure 3
Distribution of Corporate Content by Business Value



Organizations are not disposing of unneeded information in a systematic process mainly because they aren't managing their information, especially their electronic information, and therefore don't know what information to keep and what to discard. Removing unneeded or valueless information to save on storage, as well as reducing risk and cost in eDiscovery, is one of the main advantages of adopting an information governance program.

Organizations are not disposing of unneeded information in a systematic process mainly because they aren't managing their information.

An effective defensible disposal process is dependent on an effective information governance process. To know what can be deleted and when, an organization must know what information needs to be kept and for how long based on regulatory, legal and business value reasons.

ONE ADDITIONAL DRIVER: BIG DATA

Today, big data is driving changes in how information is managed, highlighting the need for true information governance. The relatively new trend to apply big data analytics to analyze huge amounts information for improved business intelligence is driving the need to better manage all enterprise information.

An information governance program addresses that need to ensure that retained information is applicable to a big data analytics program. Its policies, procedures and technologies provide a measureable and repeatable approach to managing, improving and leveraging information to increase an organization's confidence in decisions made and going forward strategies.

THE IMPORTANCE OF AUDIT IN AN INFORMATION GOVERNANCE PROGRAM

Creating policies and procedures along with new adopting new technology to ensure that all information is captured, stored, and managed appropriately is quickly

becoming a high priority for many organizations. But simply documenting policies and procedures and hoping that employees follow them do not make an information governance program. Effective information governance is crippled by poor or no employee training. For example, in the 2013 AIIM Industry Watch[vi] report only 16% of survey respondents regularly train all their staff on information governance policies and 31% do no training at all. Training the employee base is an important step in ensuring adoption and compliance with the program. But just as important is auditing the effectiveness and employee compliance with any given set of policies and procedures. As a judge once remarked at a legal conference, "an information governance policy that is not enforced is worse than not having a policy at all."

THE ROLE OF TECHNOLOGY IN INFORMATION GOVERNANCE

The amount of data organizations are dealing with today would have been inconceivable just 10 years ago. Much of this data is unstructured (greater than 80%), making it especially difficult to manage programmatically. Unstructured information in the form of email, attachments, work documents, presentations, spreadsheets and social media bury end-users day after day with very little of it actually being actively managed.

INFORMATION CATEGORIZATION

To successfully manage information for regulatory, legal and business considerations, the individual pieces of information must be differentiated from each other in some manner. What differentiates a finance record that should be retained for three years from a finance email that compares last year's finance department's overall bowling average with this year's? Proceeding to this next step in technology capability will, in reality, become a major inflection point for information governance.

The amount of data organizations are dealing with today would have been inconceivable just 10 years ago.

CUSTODIAN DRIVEN CATEGORIZATION

Because of the lack of effective tools in the past to actively manage unstructured information, organizations relied on end-users to manage their content manually: deciding what to keep, where to keep, for how long to retain it, and what to delete and when. This reliance on employees to manually categorize and file information has not been a successful practice in most organizations due to the fact that employees don't have time to interpret and actively manage the huge amounts of electronic data they deal with daily. In reality, most employees drag and drop the files they actually read to a "catch-all" folder either in their email box or on their desktop or share drive. In some minority of instances, the employee will recognize an actual "record" and move it up to a SharePoint repository with an indefinite retention period, but the vast majority of content is kept forever by default in the employee's personal storage repositories.

AUTO CATEGORIZATION

Auto-categorization employs computer programs to automatically categorize information so that employees can be free from the need to do this work manually. Computer-based categorization has made several leaps in the last couple of years utilizing keywords, rules and combinations of both. Additional advances include content- or concept-based information categorization.

- **Keyword/rules-based categorization**
  One of the obvious ways to differentiate information is by categorizing it based on simple rules. An example would be a rule to capture all content originating from employees within the Accounting department and retain it for three years. A very simple rule based on department, or based on specific keywords might be, "documents containing the word *taxes* shall be kept for a minimum of seven years and stored in the "taxes" folder on the accounting share drive" or, using a combination of rules and keywords. The main drawback with keyword and/or rules based auto-categorization is that it tends to over-retain information.

- **Content/concept-based categorization**
  A relatively new and potentially more accurate computer-based auto categorization technology is based on advanced analytics (mathematical algorithms) that enables text pattern recognition and language-context analysis to better identify the actual meaning of the content in a given document. Examples of "positive" matches are used to train the software to recognize what types of content meet the definition of the category in question. Using these advanced analytics to categorize unstructured data can help reduce the volume of data (false positives), greatly raise the accuracy level of categorization and speed the auto-categorization process.

STRUCTURED VS. UNSTRUCTURED DATA

Information is usually divided into two categories: structured and unstructured. The difference between structured and unstructured data can be simple or complex. The simple definition of the difference between the two is that structured data resides in a database while unstructured data is everything else. Digging deeper into the more complex definition, structured data resides in fixed fields within a record, such as data within a relational database. Unstructured data has no fixed fields and is also referred to as "free-form" data, such as a word document memo or PowerPoint presentation that could include rich media (sound, video, etc.).

Applications that generate structured data are mostly already managed centrally by the IT organization (except for employee-controlled applications like Microsoft Access). This is due to the fact that most enterprise databases are centrally run and managed across the entire enterprise for the benefit of the entire organization. Unstructured data usually originates from end-users running applications on their individual laptops or workstations and have direct control over where the unstructured data is stored and who has access to it. For example, Word files could be stored on a departmental share drive to which others have access, but employee use patterns and corporate culture still drives end-user unstructured data to be saved locally, meaning almost no one else knows it exists or has access to it.

An information governance program should address both types of data and include specific procedures for how end users handle their unstructured information. Otherwise, the most valuable data is lost to the organization as a whole.

THE NEED TO CONTROL HOW INFORMATION IS SHARED AND USED

One of the key reasons organizations move towards an information governance program is to enable an enterprise-wide sharing of information. This can increase the return-on-investment (ROI) of information by ensuring that valuable content is available to all that could benefit from it. An important caveat to this enterprise-wide sharing, however, is to be aware that not all information is appropriate for general consumption. For example, a work document created by the Director of HR discussing the salaries of various department head, or a document from the General Counsel outlining a current case strategy would be inappropriate for general availability across the enterprise.

An important part of an effective information governance program is the ability to designate and manage levels of information access and usage. In many cases, this can be done systematically via automation making it almost seamless and immediate (e.g., via Microsoft Active Directory). For example, tools that fully integrate with Active Directory can provide the security outlined here.

THE ROI OF INFORMATION GOVERNANCE

Calculating ROI is a performance measurement used to evaluate the productivity of an investment. ROI calculations let decision makers compare returns from various

One of the key reasons organizations move towards an information governance program is to enable an enterprise-wide sharing of information.

investment opportunities to make the best investment decision for their organization's available dollars.

ROI calculations are only as good as the variables used to calculate them. To calculate ROI for an information governance solution, the net benefit (savings or return) of an investment is divided by the cost of the investment - the result is expressed as a percentage. Information governance ROI calculations require the following data points:

- The total cost of the various activities included in the information governance processes before the investment. The primary quantifiable information governance processes are storage and storage savings, eDiscovery costs and savings, and end-user productivity gains.

- The total cost of the various information governance processes after the new solution has been installed.

- The total cost of the new information governance solution (investment).

There are four information governance activities that are relatively easy to measure before and after the adoption of an information program to effectively measure ROI. They are:

1. **Storage savings**
   Less storage purchased because a sizeable amount of data has been disposed of via accurate categorization and defensible disposition processes.

2. **Savings from storage tiering**
   Aged or less frequently accessed data is identified and moved to less costly storage (Hierarchical Storage Management, or HSM) thereby freeing up more costly storage and reducing Tier 1 or 2 storage purchase requirements in the future.

3. **Litigation/eDiscovery costs**
   The more expired or unneeded data disposed of due to effective information governance processes, the less money will be spent on eDiscovery collecting and reviewing information that should not have been kept.

4. **End-user productivity gains**
   Employees will spend less time searching for information and less time recreating content they searched for but couldn't find. The result is that both recovered time and additional revenue opportunity based on recovered hours can be estimated.

Calculating the ROI of an information governance solution can be problematic.

Calculating the ROI of an information governance solution can be problematic. Realistic information governance costs and cost savings before and after the solution is in place need to be estimated. With conservative and defendable "before and after" cost savings, an ROI for an information governance solution can be the best documentation to justify an investment in information governance.

RECOMMENDATIONS

Information governance is an enterprise-wide responsibility that can greatly add to the bottom line or end up being a complete waste of money and resources if not performed correctly. With that in mind, below are 13 common sense recommendations to ensure a successful information governance program:

- **Recommendation #1**
  Set up an information governance advisory group to ensure that every part of

the organization is represented. The group needs to agree on what they want to accomplish over a given period of time.

- **Recommendation #2**
  Establish targets, priorities and desired outcomes so that the success or failure of the information governance program can be measured over time.

- **Recommendation #3**
  With leadership from the internal legal department (and/or external counsel) fully document and understand the regulatory retention requirements to which the organization is subject (and review them annually).

- **Recommendation #4**
  Create and document the information governance program/policy/procedures, including what decision makers want to accomplish. Documenting the policy and procedures sets a baseline of understanding for the employee base, as well as a "policy of record" if it is the subject in a regulatory or legal action.

- **Recommendation #5**
  Deploy technologies to implement as much of the program as possible. In the past, most organizations relied on employees manually governing every aspect of their information – a practice that, in the long run, almost never worked. Now, with the huge amounts of digital information employees deal with on a daily basis, it's impossible for all of these processes to be manual. Reliance on employees must be taken out of the process wherever possible. For example, implementing tools that provide security and auditing capabilities, information categorization, archiving, encryption, etc. with as little employee intervention as possible will significantly improve the likelihood of success for an information governance program.

- **Recommendation #6**
  Train all employees on the new program, policy and procedures. Include the philosophy behind the program; what the organization trying to accomplish and why.

- **Recommendation #7**
  Audit and document employee compliance with the program. Again, this can be beneficial if it is the subject in a regulatory or legal action.

- **Recommendation #8**
  Enforce punishment of non-compliance with the policy. In the eyes of a court, non-enforcement equals no policy.

- **Recommendation #9**
  Target existing "dark data" repositories first for identifying sensitive or "controlled" information, such as PII, IP, or regulatory content for migration to more secure managed locations. In many cases, these huge data repositories contain high levels of eDiscovery and regulatory risk, as well as potential cost reduction.

- **Recommendation #10**
  Target the same "dark data" repositories with a defensible disposal process targeting duplicates, old and valueless content. In many cases, defensible disposal processes can clear 40-50% of data on a file share, thereby reducing eDiscovery and regulatory risk and storage cost.

- **Recommendation #11**
  Measure the information governance program results: storage savings, eDiscovery and productivity.

Set up an information governance advisory group to ensure that every part of the organization is represented.

- **Recommendation #12**
  Document everything, including, policies, procedures, program goals, employee training, audit practices, enforcement actions and program outcomes. This practice helps in tracking progress, as well as outcomes. Documentation is always helpful when dealing with a regulatory or legal action, as well.

- **Recommendation #13**
  If the legal department permits it, publish the results, success or failure of the information governance program to help others in defining their information governance programs.

SUMMARY

As has been stated several times in this white paper, organizations run on information and successful organizations are more successful when they actively manage all of their information. In these times of dramatically expanding information stores, out-of-control eDiscovery costs, and increasing regulatory requirements, ignoring the risks of unmanaged information is a losing and costly strategy. Sitting on huge amounts of unmanaged dark data is a competitive disadvantage when competitors are effectively managing and using all of their information assets.

Successful information governance programs combine records management, archiving, eDiscovery processes, technology-based auto-classification, and other capabilities to help organizations reduce legal and business risk, and to drive business value. When information is actively managed making it easier to find, use, manage, and dispose of, the organization profits greatly.

SPONSORS OF THIS WHITE PAPER

At Actiance, we've been helping our customers manage new forms of communication since 2002. We work closely with regulators including the SEC, FINRA, IIROC, and the PRA and FCA, and with our customers, to ensure that they understand the capabilities of today's technology and that our platform meets their most stringent requirements.

This rich heritage and extensive experience uniquely positions Actiance to address our customers' information governance challenges.

We enable your organization to use the Unified Communication, Enterprise Social Software, Instant Messaging, social networking, and custom-built enterprise apps it needs with the regulatory, legal, and corporate compliance it requires.

Our platform is a foundation you can build on now and in the future. It integrates with your existing infrastructure and scales as your requirements change so you can unleash social business.

Actiance is based in Silicon Valley and operates offices in North America, the United Kingdom, and India. Actiance is privately held with funding provided by Credit Suisse, JK&B Capital, Scale Venture Partners, and Sutter Hill Ventures.

actiance®

www.actiance.com

@Actiance

+1 888 349 3223

info@actiance.com

ownCloud helps enterprises concerned about sensitive data leakage via Dropbox deliver a secure file sync and share solution on site, on their storage, integrated with their infrastructure and security systems, managed to their policies. The result is an easy-to-use solution that provides complete control over sensitive corporate data.

ownCloud integrates seamlessly into existing user directories, governance, security, monitoring, storage and back-up tools — becoming part of the existing infrastructure. And because ownCloud is open source and open by nature, plug-in apps exist to extend ownCloud out of the box, enabling LDAP/AD integration, file versioning, file sharing, external file system mounts and much more. If an application or capability needed is not there, simply create a new application plug-in and add it to the ownCloud server.

For more information visit www.owncloud.com.

www.owncloud.com
@ownCloud
facebook.com/owncloud
+1 877 394 2030

Rand Secure Data, a division of Rand Worldwide (OTCBB: RWWI), is a leading provider of data governance solutions that combine ultra-security and high-performance with simplicity, accessibility and affordability. Rand Secure Data offers industry leading data archiving, email archiving, backup and eDiscovery solutions that allow companies to simplify IT strategies, lower overall costs and free up resources without having to make any sacrifices in terms of functionality, security or control.

As a division of Rand Worldwide, one of the world's leading professional services and technology companies for the engineering community, Rand Secure Data has more than 30 years of experience developing and delivering technology solutions to customers and has offices throughout the United States and Canada. To learn more visit www.RandSecureData.com.

www.randsecuredata.com
@randsecuredata
facebook.com/randsecuredata
+1 877 443 8042
RandSD@rand.com

Storage Made Easy provides an enterprise file share and sync cloud control software appliance that can be deployed on-site, in a trusted data center or on IaaS infrastructure such as Amazon Elastic Compute.

The SME solution allows IT to regain control of "cloud data sprawl" by unifying private / public company data and privately used employee cloud data solutions into a single converged infrastructure. This can easily be managed and be used to set and enforce governance and audit controls for file access and sharing in addition to providing deep content search of indexed data.

This approach provides a solution to the "shadow IT" conundrum and makes it possible for companies to find a balance between the protection of corporate data and employee data by allowing businesses to monitor, secure and audit all data silos, be they private or cloud / company or employee, from a single access point.

www.storagemadeeasy.com
@SMEStorage
facebook.com/ StorageMadeEasy
Skype: storagemadeeasy
sales@storagemadeasy.com

REFERENCES

[i] AIIM Industry Watch: Information Governance – records, risks and retention in the litigation age

[ii] RAND Institute for Civil Justice Report "Where the Money Goes"

[iii] *The Growing Need for Mobile Device Archiving,* Osterman Research, Inc.

[iv] http://www.serverwatch.com/server-trends/nirvanix-shut-down-sends-shockwaves-through-the-cloud-services-industry.html

[v] http://www.usatoday.com/story/news/nation/2013/12/18/secret-service-target-data-breach/4119337/

[vi] AIIM Industry Watch: Information Governance – records, risks and retention in the litigation age