



## A Practical Enterprise Methodology for Addressing the Compliance Challenges of eDiscovery, eRetention Management, and Defensible Disposition

---

# A Practical Enterprise Methodology for Addressing the Compliance Challenges of eDiscovery, eRetention Management, and Defensible Disposition<sup>1</sup>

Andrew M. Cohen, Esq.

EMC Corporation

Associate General Counsel, and

Global Solutions Practice Lead, Compliance<sup>2</sup>

## Section I: Introduction and Executive Summary

### 1. Overview

The purpose of this white paper is to articulate a practical framework for understanding and proactively addressing eDiscovery and eRetention management compliance challenges. With the explosion of electronic information, organizations are facing escalating costs of legal discovery—which can be as much as a million dollars or more for a single case—as well as the challenges and risks associated with retaining (and disposing of) electronic information. While the core goals of eRetention management (i.e., keep information while it has business and/or legal value, then get rid of it) and eDiscovery (i.e., efficiently find and produce information responsive to lawsuits and government investigations) are relatively simple to state, for many, they have proven difficult to achieve.

Many organizations struggle to address these needs because they have difficulty bridging the gap between the Legal, Records Management, and IT functions, because they don't have proactive control over their information infrastructures, and because they lack a practical roadmap for where to begin. This paper seeks to provide such a framework, and proposes that organizations initiate an enterprise strategy broken into digestible pieces, based on business and legal priorities, risk management, repeatable process, and return on investment ("ROI"). The major sections of this paper are as follows:

- **Section I.** Provides an introduction to **key concepts**, and an **executive summary** of the ten-step methodology proposed by this paper.
- **Section II.** Explains and summarizes key **pain points and current legal and technology trends** related to eRetention management and eDiscovery.
- **Section III.** Explains the ten-step **methodology** in detail.
- **Section IV.** Addresses management issues specific to **e-mail**.
- **Section V.** Conclusion.

<sup>1</sup> This white paper is not a legal opinion and it does not constitute legal advice. For all questions regarding compliance with statutes, regulations, procedural rules, or other legal obligations, please seek legal advice.

Entire contents copyright 2006 EMC Corporation. Reproduction of this publication in any form without prior written permission of EMC Corporation is forbidden. EMC Corporation shall have no liability for errors, omissions, and the reader assumes sole responsibility for the selection of these matters to achieve intended results. The opinions expressed herein are subject to change without notice. All rights reserved.

<sup>2</sup> Mr. Cohen has a unique set of responsibilities at EMC Corporation. In his legal capacity, Mr. Cohen and his team manage EMC litigation, investigations, employment issues, and aspects of corporate compliance including training and eRetention management. In his business role, Mr. Cohen and his Compliance Solutions Practice team are driving EMC's delivery of integrated compliance solutions, including for proactive eDiscovery (services, software, and hardware). Mr. Cohen is an active member of the Sedona Conference, and he is a frequent speaker and writer in the areas of eDiscovery, e-mail archiving, content management, and eRetention management.

---

## 2. Today's Compliance Challenges—Information Management and Bridging the Gap

Compliance is a complex and malleable concept. However, stated simply, compliance means identifying criteria with which an organization will comply (e.g., laws, regulations, industry standards, rules, policies, and other sources of mandatory or voluntary criteria), and building processes and infrastructure to achieve and prove compliance with those criteria.

In today's organizations, electronic information is at the heart of compliance. Take, for example, the e-mail message containing a discussion thread between executives that is relevant to a lawsuit or regulatory investigation; the final "signed" electronic contract; the online workflow that evidences the steps an organization took to develop a product or service, such as a new drug or an insurance plan; the audit trail records that show that an internal control is being followed for purposes of Sarbanes-Oxley; the log file that shows that access controls were in place and that a customer's financial data was kept secure; and so on.

For many organizations, managing electronic information in a compliant manner has become a central concern—and rightfully so. The ongoing penetration of information technology into the enterprise enables tremendous efficiencies—but also generates unprecedented volumes of information that must be properly managed, retained, and discovered. Why has this aspect of compliance become so painful, so confusing, so risky, and so expensive? In large part, because Legal and Records Management processes, and IT infrastructures and processes have not kept up with the exploding velocity of information creation, especially for unstructured e-documents and e-mails.<sup>3</sup> At the same time, legislatures, courts, and regulators have been demanding that organizations be transparent not only with financial data, but also with general business communications (which today are found in e-mail, instant messages, e-documents, voice-over-IP, blogs, and so on). Governments are also requiring greater privacy and security, especially for customer and employee personal financial and medical information.

Traditionally, information management and legal discovery were likely to have been addressed as separate issues, by separate departments. Today, however, there is a need to bridge this gap by bringing Legal<sup>4</sup> and IT together to address the compliance challenge. *The lawyers, compliance officers, and records managers increasingly need to understand where the electronic information sits and how it is technologically stored and managed. The IT folks are being asked not just to "back up" information, but also to help classify it, policy manage it, and efficiently search, retrieve, and produce it.*

The core business need to manage all this information, together with the dramatic costs, legal risks, and the inherent desire of good organizations to meet their legal obligations, have caused the somewhat dusty concepts of records retention, legal discovery, archiving, and content management to converge and to move dramatically up the corporate stack. Organizations have an acute and strategic need to proactively manage information, to set and enforce policies that also are practical (which often includes being low impact on busy employees), to extract more value from information, to stop wasting money managing, duplicating, and discovering it, to dispose of it in a defensible way, and to avoid showing up on the front page of the *Wall Street Journal*.

<sup>3</sup> **Structured information** is what is in databases. For example, a system for tracking airline reservations, or a financial reporting system. **Unstructured** information is essentially e-documents (such as word documents, spreadsheets, presentation slides, pdfs, and so on). **Semi-structured** content is e-mail. From a technical perspective, e-mail is organized into a series of fields—the "to" field, the "from" field, the "body" field, and so on. In that sense, it has some structure, but e-mails are objects that are not part of an integrated database, and e-mail often sits unmanaged in a variety of places in the enterprise.

<sup>4</sup> For simplicity, in this paper, the term "Legal" may collectively include Records Management, Compliance, Internal Audit, and any other compliance-related function.

---

### 3. The Convergence of Records Retention and Legal Discovery

Records retention and legal discovery were traditionally handled as two separate domains. Records retention involved a paper-based process of reviewing documents, page-by-page, declaring certain documents as official “records,” classifying them by comparing them to a long manual of retention periods, and putting the documents into physical storage with the associated retention periods. Those documents that were not declared official “records,” such as drafts or routine communications, were not supposed to have been retained. Depending on the organization, records management was handled either by professional records managers and/or by various individuals across an organization, such as administrators.

In contrast, legal discovery involved a separate process of lawyers responding to subpoenas or other discovery requests in government investigations and lawsuits. The lawyers would essentially identify potential witnesses and ask them to print out or collect documents that might be relevant. Then the lawyers would review those documents and provide the appropriate materials to the other side in the case.

Today, these two disciplines are converging, both from a legal process perspective and from a technology perspective, because they are struggling to address the same issue—the explosion of electronic information. Whether the question is how long to keep certain types of information (“eRetention”), or how such information is going to be collected, preserved, and produced for litigation (“eDiscovery”), there is an acute need for repeatable and cross-functional business processes and an IT infrastructure that allows proactive information lifecycle management (“ILM”). ILM is the cradle-to-grave policy management of information based on its content and changing value over time. In other words, organizations need to establish appropriate policies, together with enterprise systems that allow them to efficiently manage information while it has business and/or legal value, and then to defensibly dispose of it, and those same systems should have the built-in capability to search, collect, hold, cull, and produce information for eDiscovery.

### 4. Executive Summary: Methodology

This paper proposes a ten-step methodology for addressing organizational eDiscovery and eRetention management compliance challenges. This methodology, which is summarized below, expressly does not seek to solve every problem associated with these challenges. Rather, it is designed to provide a high-level, practical framework that enables organizations to simplify some of the issues, establish priorities, and get started on addressing the key challenges. This methodology is explored in more detail in Section III.

1. **Act cross-functionally.** Identify the right people, and take a cross-functional approach to eDiscovery and eRetention management compliance. In many organizations, IT, Legal, and others are in fact all struggling to address the same issues from their own unique vantage points. Alone, IT often needs assistance to determine what the legal and business policies should be, and Legal often does not fully understand how the information is managed, and consequently, does not fully understand the risks and costs of various policy alternatives. Given the amount of money being wasted, and risk being incurred, cross-functional communication is a logical and necessary initial step.
2. **Perfection is the enemy of the good, so use risk management.** No organization has achieved perfect information management compliance. A compliance person’s “ivory tower” vision of the perfect retention policy—which gets ignored in practice because employees are too busy to classify the 200 e-mails they receive each day—ends up

---

being counterproductive. In this context, perfection is the enemy of the good. What makes sense is to discreetly but honestly assess your organization's key gaps, establish priorities based on where the most risk is being incurred and the most money is being wasted, and start with a realistic plan to close those gaps.

3. **Assess the impact on employees.** It is important to understand the impact on employees of different approaches to eRetention Management. As you assess process and technology solution alternatives, recognize that changing employee behavior to achieve a compliance goal can be difficult. Changing employee behavior may well be justified in some circumstances, but you should consider "picking your spots" when it comes to how much you impact busy employees in their day-to-day activities.
4. **Build a top-down understanding of your information assets.** Take a top-down approach to understanding the types of information your organization creates and where it resides. Since the vast majority of information spends its entire lifecycle in electronic form, it makes sense to inventory the "big buckets" of information in your organization (especially by application), rather than trying to catalogue every individual document that exists in an enterprise. Often, useful collateral already exists within IT, such as lists of company applications and systems inventories. The top-down approach will help you get a grasp of the information assets in your organization in a relatively short period of time.
5. **Prioritize and don't "boil the ocean."** Once you have conducted diligence top down, you can prioritize and start with the activities that will address the most acute pain points and/or provide the greatest return from a legal and/or business perspective.
6. **Leverage specialized consulting resources.** If appropriate, leverage specialized and focused consulting resources that can help bridge the gap between Legal and IT by conducting cross functional interviews and workshops, identifying priorities and key gaps, helping to build a business case, and providing expertise to identify and address areas of significant risk.
7. **Simplify existing retention policies and incorporate reasonable, repeatable, cross-functional business process for retention, disposition, and discovery.** Many companies have record retention manuals containing sometimes hundreds of discrete retention periods (that were originally meant to be applied to paper documents). The sheer volume of electronic content makes these manuals difficult if not impossible to apply to the electronic content that forms the vast majority of organizations' information assets. The complexity of paper-focused retention schedules should be dramatically reduced so that policies can begin to be operationalized and applied automatically to electronic content, by application, including at the time the information is created. As discussed further below, the right repeatable processes, including for eDiscovery and litigation hold, allow a company to defensibly dispose of content at the end of its lifecycle.
8. **Build a simple business case especially where there is a cross-functional return on investment (ROI).** Develop a high-level business case that evaluates the economic impact of planned changes to information management. When considering archive and content management solutions, assess the cost savings associated with de-duplication and more efficient information management ("IT ROI"), as well as the cost savings from more efficient data collections for eDiscovery ("Legal ROI"). A simple business case is often not difficult to construct.
9. **Implement an enterprise ILM strategy broken into digestible pieces, consider a "big-buckets" and then a "little-buckets" strategy, and build proactive eDiscovery efficiencies into the IT infrastructure.** Make information technology decisions carefully and choose a scalable infrastructure that manages information based on its changing value over time,

---

rather than point solutions that will leave stove-piped information stores. Break down these projects based on priority. For many organizations, an appropriate initial technology step involves “cleaning up” the infrastructure by establishing central archives and content management repositories (“big buckets”), especially for e-mail and unstructured e-documents that are sitting unmanaged on backup tapes, servers, and employee hard drives, that are being duplicated over and over again, and that cannot be policy-managed or easily searched for eDiscovery. Over time, solutions can be built into the organization’s same IT infrastructure that allow for increasingly granular eRetention policy management (“little buckets”), and the ability to automate searches across different data repositories, conduct data collections, legal holds, and culling for proactive eDiscovery.

10. **Continuous improvement.** Conduct regular reviews, audits, training, and cross-functional oversight to refine processes, and apply technology improvements when they make sense.

## Section II: Pain Points and Challenges

A number of recent technology and legal developments have combined to make eRetention management and eDiscovery expensive and risky for organizations.

### 1. Records Management Paradigm Shift

Records management is a programmatic effort by an organization to classify and policy-manage its information in a proactive way. At most organizations, records management programs are still paper-focused, supported by detailed retention schedules or manuals that contain hundreds of record types and retention periods—with an expectation that each record will be identified and classified by a human being. Increasingly, this approach does not support the realities of today’s digital business environment.

Electronic content is different than paper content. The volume of electronic content is dramatically greater, it can be duplicated and transported much faster, it contains metadata, and so on. Paper is still relevant, but over 90 percent of information is created electronically and will stay in an electronic form throughout its lifecycle.<sup>5</sup> Approximately 30 billion e-mails are sent *every day*,<sup>6</sup> and a typical 40-gigabyte laptop hard drive can hold the equivalent of 20 million pages of information. Despite these dramatic changes, most organizations continue to try to apply paper-based records management processes. Given the velocity of information generation, the old method of page-by-page manual document review, classification, and declaration of each document as a “record” or not a record alone cannot work.<sup>7</sup> As discussed in Section III, a new paradigm is necessary to address the proactive management of electronic content.

<sup>5</sup> How Much Information 2003” by UC Berkeley’s School of Information Management and Systems (2003).

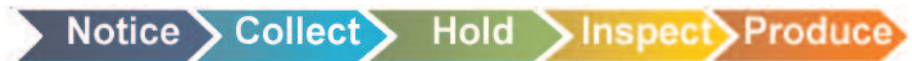
<sup>6</sup> Worldwide E-mail Usage 2005-2009 Forecast: E-mail’s Future Depends on Keeping Its Value High and Its Cost Low” by Mark Levitt and Robert P. Mahowald, IDC (December 2005).

<sup>7</sup> 2005 Electronic Records Management Survey: A Call to Action” by Robert F. Williams and Lori J. Ashley, Cohasset Associates, Inc. (2005).

---

## 2. eDiscovery—The Massive Volumes versus Inadvertent Destruction Dilemma

Electronic discovery is the legal obligation of organizations to produce information that is or may be relevant to the subject matter of a lawsuit or government investigation. Today, eDiscovery is imposing significant burden, risk, and cost on organizations. At a high level, eDiscovery involves the following simplified process:



In essentially every case, there is some notice of a new obligation (such as a new complaint or a government subpoena), an effort to collect relevant information for the case, a hold or preservation of that relevant content to ensure that it is not destroyed prior to when it needs to be produced, a review of the collected content by attorneys (including for relevance and privilege), and ultimately, the production of the relevant information to the other side. At each step in the process today, there is tremendous cost, waste, and risk. Think of eDiscovery as a sometimes gigantic culling exercise, where initial collections start broadly and are refined; the more information that is collected initially and the more manual the exercise, the greater the costs. Manual and reactive collections of information lead to significant costs for processing data to get it into a form that can be searched and reviewed by attorneys, and the more information that gets sent to attorneys, the more money that is spent for hourly attorney reviews.

The eDiscovery challenge raises “*the massive volumes versus inadvertent destruction dilemma*.” Organizations are piling up massive amounts of information, which creates costs and management challenges for the IT department, but also drives growing *legal* costs and risks. The greater the universe of unmanaged content, the more money is spent collecting, processing, reviewing, and producing it for eDiscovery.

On the other side of the dilemma, organizations face legal and reputational risks associated with destroying information and later being second-guessed and accused of *spoliation*, or the wrongful destruction of evidence.

Organizations are struggling to determine how to maneuver between these two competing concerns. The answer, in short, is proactive information management and a repeatable cross-functional business process for eDiscovery.

### 2.1 Massive Volumes

The piling up of unmanaged information results in two key eDiscovery-related cost drivers:

1. Increasing costs of reactively searching, securing, preserving, and processing the data collected.
2. Increasing costs of attorney review.

At most organizations, electronic information sits unmanaged on employee computer hard drives, networked shared drives, and on backup tapes. Collection of such unstructured and semi-structured information (see footnote 3) for eDiscovery purposes sometimes requires the mirroring of all the contents of witness hard drives, collection from shared drives, and/or all the contents of various backup tapes by server. Since there is no automated way to effectively narrow the universe of the initial information collection, massive amounts are collected, most of which are completely irrelevant to the case.

Organizations have to expend significant internal IT resources or use reactive eDiscovery vendors, who manually collect such information, process it into specialized litigation repositories, and host those repositories. Vendors in the traditional eDiscovery space

---

can provide valuable services. *However, hundreds of thousands or even millions of dollars can be expended on data collection, processing, and monthly hosting fees for a single case, and when that case is over there may be nothing to show for it.* Those dollars are gone, the company's IT infrastructure has not been improved, and when the next new large case hits, a new set of reactive data collection dollars go out the door.<sup>8</sup>

*With no efficient way to cull down the collections, organizations have little choice but to take the huge universe of information and "throw it over the wall" to their outside counsel, who in turn charge significant hourly rates for teams of associates to review the content.* Outside attorneys may well have good intentions for efficiently reviewing content, and they may even have access to technology tools to cull down the universe of what has been collected and placed into a litigation repository. However, they also have strong incentives to review much of what is provided by the client, because they are paid by the hour, and it might be viewed as malpractice to fail to review an important document or piece of information that the client had turned over. Indeed, a recent National Law Journal article reported that, due to the dramatic increase in information subject to eDiscovery, law firms are hiring unprecedented numbers of temporary employees to engage in these brute force document reviews.<sup>9</sup>

## 2.2 Inadvertent Destruction—"Save Everything"?

When a big lawsuit or government investigation hits, it is not uncommon for someone in the legal department to call the IT department and instruct them to "save everything." Indeed, some organizations have declared that, because they don't know exactly which content is subject to eDiscovery, and because they have a number of active litigation and investigation matters at any given time, their company's "strategy" is to place everything on litigation hold and therefore save everything. Despite the costs and risks of piling up massive volumes, many organizations continue to do so because they don't have control over the information in their enterprise and they are more concerned about the legal and reputational risk of destroying the wrong thing. *For most organizations, there is a need to gain proactive control over their information since saving everything forever is not going to be a sustainable strategy.*

## 3. E-Discovery Legal Trends

Current legal trends underscore that the risks on both sides of the dilemma—either for piling up more information, or for destroying it the wrong way—continue to increase.

### 3.1 The New Federal Rules of Civil Procedure

The Federal Rules of Civil Procedure are going to be amended, likely by the end of 2006, to include special provisions on eDiscovery.<sup>10</sup> On the one hand, these provisions should provide organizations with some limited protection if the company acted in good faith but still negligently allowed relevant content to be destroyed as a result of routine operation of systems. But on the other hand, the new rules will make eDiscovery a focus of every federal case in the United States.<sup>11</sup> *It is noteworthy, that under the new rules, it will increasingly become a best practice for organizations to have an inventory of information "sources" and the ability to identify from which sources they are, or are not, producing information.* Other countries are also following the lead of the United States in this area.<sup>12</sup>

<sup>8</sup> Often organizations are not fully aware in the aggregate of how much money is being spent on these reactive services, since payments to eDiscovery vendors are often subsumed in individual invoices from outside counsel, who bill separately for each case/matter that they handle.

<sup>9</sup> "Rising Tide of E-discovery" by Sue Reisinger, National Law Journal, Volume 27, Issue 53 (September 19, 2005).

<sup>10</sup> See e.g., "Playing by the Rules: The impact of the new E-discovery Rules" by Thomas Allman, Mayer, Brown, Rowe & Maw LLP. (March 2006).

<sup>11</sup> Attorneys and investigators have already begun to realize that they can leverage the costs and pain of eDiscovery for strategic advantage in a case. For example, rather than beginning a case by deposing the witnesses who have knowledge of the facts, lawyers sometimes start a case by taking the "corporate representative" deposition of an internal IT person, to learn about the IT environment, and how data is stored and destroyed. This information may then be used to create leverage in the case, especially if the company does not have proactive control over its information.

---

### 3.2 The Coleman (Parent) Holdings Case

In a case that demonstrated the potential minefields hidden in eDiscovery,<sup>13</sup> plaintiff Coleman Holdings alleged fraud in connection with a corporate merger. The defendant, Morgan Stanley, vigorously denied the claims. Coleman pushed for massive discovery of company e-mails, which defendants kept on tapes in various warehouses and other locations. The collection, review, and production of content from those e-mail backup tapes were extremely burdensome and difficult. The plaintiff successfully argued that the defendant had systematically failed to produce information that was subject to discovery. Eventually, the judge ruled that defendant's eDiscovery violations justified an adverse inference on the underlying fraud claim. Essentially, the judge held that the defendant's inability to produce e-mails demonstrated that the underlying fraud allegations were in fact true. The company, because of process and technology breakdowns and its inability to produce the e-mails, never really got to defend its position on the merits. The jury returned a verdict of over *one billion dollars* against the company. Although an extreme example, this case underscores the potentially massive costs and risks associated with trying to conduct discovery from a warehouse full of e-mail backup tapes, and the failure to have an effective cross-functional discovery process.

### 3.3 The Zubulake Case

In what was otherwise a routine employment case, the jury awarded the plaintiff \$29 million. The significance of this case, however, is contained in Judge Scheindlin's series of written decisions, especially the way that she addressed the obligation of organizations to effectuate "*litigation holds*."<sup>14</sup> A litigation hold is the identification of information that might reasonably be relevant to a case, and the preservation of that information (including beyond its normal retention period if necessary) to make sure that it is available for future production to the adversary in a case. The strength of the reasoning in the Zubulake decision has had significant impact on the law, but the case has also created huge concerns for organizations since it articulated an extra layer of responsibility beyond the obligation to set reasonable policies and follow them.

By way of example, assume that an organization with a blanket one year e-mail retention period receives notice that a new lawsuit has been filed on January 1, 2005. But the opponent's request for production of documents is not sent to the company until 18 months later, on June 30, 2006. In the past, the company might not have systematically collected and preserved documents until after the June 30, 2006 document request was received. Indeed, the company might have openly acknowledged that certain e-mails could no longer be produced as of 2006 because they were destroyed pursuant to the company's normal one year retention period. In Zubulake, Judge Scheindlin held that such a position would not be sufficient. Rather, the company would be obligated, within a reasonable time after it received notice of the case on January 1, 2005, to take reasonable steps to collect information that was relevant to the case, and to preserve that information until it is subject to production, even if that is beyond its normal one-year retention period.<sup>15</sup> Clearly, for many organizations, the costs, complexities, and risks associated with trying to manage eDiscoveries and litigation holds across an enterprise requires repeatable processes and proactive and more automated technology solutions for eDiscovery.

<sup>12</sup> For example, both the UK and Canada issued new eDiscovery guidelines in 2005.

<sup>13</sup> Coleman (Parent) Holdings, Inc. v. Morgan Stanley & Co., Inc., 2005 WL 679071 (Fla. Cir. Ct. Mar 1, 2005).

<sup>14</sup> Zubulake v. UBS Warburg, 2004 U.S. Dist. LEXIS 13574 (2004) ("Zubulake V").

<sup>15</sup> Ibid.

---

### 3.4 Sarbanes-Oxley

Section 802 of the statute makes it illegal for any person to knowingly alter, destroy, or conceal information with the intent to impede or obstruct any federal government investigation. Violations are punishable by up to 20 years in prison. 18 U.S.C. §§1512, 1520 (2002). Good organizations know that shredding, or intentionally destroying relevant documents is illegal and wrong, but Sarbanes has raised the stakes given the magnitude of the penalties and the sweeping language of the statute<sup>16</sup>.

### 4. Why E-mail and Unstructured E-Documents are a Priority

Improving e-mail management is a focus and priority for many corporations because e-mail is currently where the eDiscovery and eRetention management pain points meet in an acute way. First, from a retention management perspective, e-mail is a huge challenge because “e-mail” is not a type of record. Rather, e-mail is a messaging system containing a huge range of content—from critical content such as contracts, trade secrets, and regulatory materials on the one hand, to content that has no value or negative value to the organization, such as idle chatter or dirty jokes, on the other. Second, e-mail is often the focus of eDiscovery requests, and when it is unmanaged, it creates significant costs for production, and risks associated with litigation hold and inadvertent destruction. Third, e-mail is typically a large, enterprise-wide application, which can be expensive to manage. E-mail systems were originally built as a delivery mechanism for simple messages, but for many organizations, e-mail has become the way business is done, and it contains key workflows and repositories of critical business information which are unmanaged and growing exponentially.

Consider the issue of “.pst” files (in Microsoft Exchange environments), or personal archives. In an effort to control the exploding volumes of information in the e-mail environment, many organizations put e-mail box-size limitations on their users, hoping that when the size limit was reached, the users would delete unnecessary e-mails. Instead, the law of unintended consequences prevailed. Users routinely took huge volumes of e-mails and copied them in bulk to personal archives sitting on their hard drives, or to networked shared drives. These personal archives have proven to be quite costly to companies. First, because each user created his or her own personal archive, there was no true corporate e-mail retention policy (or at least no policy that was in fact being enforced). Often, in practice, no e-mail content ever gets deleted from the .psts. Second, when a discovery request hits, it is very expensive to collect such e-mails from individual archives, as this typically involves significant processing and attorney review costs. Third, from a business perspective, information that is managed by individuals on their own “personal” archives is often not sufficiently shared within the organization. Fourth, from an IT perspective, personal archives drive further duplication and backup challenges.<sup>17</sup>

There is essentially no automated and enforceable way to achieve eRetention management, and defensible disposition, of e-mails and other unstructured content when it sits in so many places. For example, when an employee generates an important PowerPoint presentation and sends it by e-mail to several dozen people within an organization, it can be duplicated on numerous servers and employee hard drives. In the absence of an

<sup>16</sup> Prior to Sarbanes, it was unclear when the obligation to preserve was triggered, but it was thought to be after an actual subpoena was received (“The Sarbanes-Oxley Act: Understanding the Implications for Information and Records Management” by Randolph Kahn and Barclay T. Blair, Kahn Consulting, Inc.). After Sarbanes, all organizations were placed on a kind of general notice that if they have anything that could be relevant to the subject matter of a federal investigation, they must preserve it (even if no subpoena has yet been issued) (SEC. 802. CRIMINAL PENALTIES FOR ALTERING DOCUMENTS – Sarbanes-Oxley Act of 2002.).

<sup>17</sup> Personal archives create a dilemma regarding disaster recovery backups for user hard drives. If an organization chooses not to back up such information (with a remote snapshot) and a hard drive crashes, valuable information could be lost. Alternatively, if the company does choose to back up the hard drives, then there is a cost to maintaining the backup environment, and unlike archives, such disaster recovery applications do not have robust policy management and search capabilities; they’re merely intended to allow the point-in-time copy of a crashed hard drive to be copied back on to a new hard drive.

---

archive, and depending on the organization's backup strategy, that exact same document may literally be backed up to multiple tapes from multiple servers, *every night*. Consider the information management impact of this in enterprises where a million (or many millions of) e-mails are sent and received *every day*, which is not uncommon. With an archive, that same PowerPoint presentation could be saved as one single object on a central repository, with multiple pointers to it, and a retention and disposition policy automatically applied and enforced. The de-duplication and proactive information management benefits of this are apparent. Moreover, if that same document were subject to discovery, the costs of collection out of an archive may be negligible, whereas from tape, the costs could be massive.

## 5. Disaster Recovery Backup, Archive, Content Management, and Federated Search Defined

It is helpful to understand at a high level the types of proactive information management solutions that are available, and how they fit together.

There is a critical distinction between “disaster recovery backup” and “archive.” *Disaster recovery backup*, for example in the form of e-mail backup tapes, is intended to be a point-in-time copy of information that can be available in the event of a disaster. For example, if a server were to stop functioning, a backup tape could be used to reload a snapshot of last week's data onto a new server.

Archives, for purposes of this paper, are defined as online repositories, where information can be automatically stored in a centralized way, indexed, searched, and retrieved in an automated way (such as by key word), duplicates can be eliminated (*de-duplication*), and where data can be classified, and automated retention policies can be applied and enforced. Many organizations have collected warehouses full of tapes and used them as a kind of pseudo archives (thus mixing their backup and archive strategies). However, because tapes are not indexed, it is very difficult to find specific content on tapes, making eRetention policy enforcement difficult if not impossible and eDiscovery expensive and risky. Archives can be created for almost any kind of electronic content. The most cutting-edge solutions offer archive technology that is integrated with content management and federated search, for even greater functionality.

*Content Management* is a technology solution that allows organizations to put essentially any kind of content into a structured environment. This allows information within the content management environment, among other things, to be policy-managed and searched. Moreover, content management allows for the check-in/checkout, version control, security, and library services for business documents. Content management also allows for the review, revision, approval, and publication of various content types as part of a business workflow. For example, content management might be used to automate the step-by-step workflow associated with getting FDA approval for a new drug. *Enterprise Content Management (ECM)* is the capability of putting a content management structure around all unstructured content in the enterprise (Web content, desktop, supply chain content, e-mail, fax, IM, scanned paper images, and so on). ECM includes: (i) Web content management (WCM) for automating the webmaster bottleneck, and managing dynamic content and user interaction; (ii) records management for long-term archiving, automation of retention and compliance policies, and ensuring legal and regulatory record compliance; (iii) document capture and document imaging for capturing and managing paper documents; (iv) document-centric collaboration for document sharing and supporting project teams; and (v) workflow for supporting business processes, routing content, assigning work tasks and states, and creating audit trails.

---

*Federated Search* is the capability of automating the collection of content within an enterprise with a tool that allows the search and retrieval of content from various repositories within an enterprise. For example, an organization might have Web content, an e-mail environment, multiple document management repositories, and so on, all in different applications. Federated search tools allow you to “reach into” these various applications and view or collect content by key word and other intelligent search. The remote policy management of such content is sometimes referred to as *Virtual Records Management*.

## **Section III: A Practical Methodology for eDiscovery, eRetention Management, and Defensible Disposition**

This section outlines ten core principles that are designed to enable organizations to prioritize their eDiscovery and eRetention management compliance activities so that they can begin to build an effective management program.

### **1. Act Cross-Functionally**

Cross-functional involvement and representation is critical to the success of any eDiscovery and eRetention management compliance program, and of course, you need the right people involved. This typically means people who are able to see the big picture and who are willing to examine the issues from multiple perspectives. In the past, it might have been hard to get cross-functional interest, but at most organizations, the costs and risks are large enough that people across functions are struggling to achieve more efficient and defensible information management. In other words, different people across different functions are often trying to address the same problem. They’re just doing so from their own vantage point. Someone just has to find them and bring them together.

At most companies, there are at least three constituencies that should participate and be represented in such a program: legal/records management/compliance/internal audit (collectively “Legal”), Information Technology (“IT”), and employees/users/lines of business (collectively “Employees”). To summarize and simplify the interests of each:

- Legal wants appropriate policies, and supporting processes, that allow for the retention and defensible disposition of information once it no longer has business or legal value, and increased automation of data collections and litigation hold for eDiscovery;
- IT wants cost reduction, improved information management, ease of administration of technology solutions, and direction from Legal on what the policies should be; and
- Employees want technology that makes them more efficient, they want access to information, and they don’t want to be bothered by compliance activities.

For many organizations, it is logical to establish a cross-functional oversight group to monitor and align the activities of different functions, manage a strategic program, and to be a forum to establish policies. At some organizations, this group already exists. At others, this group would have to be formed. It sometimes makes sense to keep the oversight group relatively small, and to staff it with senior-level people, who are in a position to establish policies and drive decision-making.

### **2. Engage in Risk Management; Perfection is the Enemy of the Good**

Addressing eRetention management and eDiscovery challenges involves risk management. No company has achieved perfect information compliance, and most have significant gaps. The key is to take steps to understand those gaps so that the organization is in a position to address the biggest priorities based on where there is the greatest need for better information management, the most risk, and where the most money is being wasted.

---

Consequently, it is important to gain a discreet but honest assessment of the current state of the company's compliance, and focus on policies and solutions that will help close the priority "gaps." For example, it may make little sense to declare that your e-mail policy is 90 days if in fact everyone at the company can (and does) create personal archives on their hard drives, allowing e-mails to be kept indefinitely. The existence of that policy will not protect the company from the reality that those e-mails exist and that they are an information "source" subject to production in an eDiscovery.

Similarly, there is sometimes a tendency to compare a proposed information management solution to a hypothetical perfect world, conclude that the solution doesn't do enough, and therefore do nothing as a result. However, with unmanaged information exploding, doing nothing may be the wrong choice. In assessing information compliance solutions, the right question is: "is this a tangible and strategic improvement" and not "is this perfect." Perfection is the enemy of the good when it comes to driving eRetention management and eDiscovery solutions.

### 3. Assess the Impact on Employees

It is also critical—up front—to embrace the reality of how employees are likely to respond to a chosen eRetention management policy. If a new policy requires employees to change the way they do their jobs, they will often try to find a way around it or ignore it, especially if it does not deliver business efficiency. For example, some organizations merely declare that e-documents including e-mails are subject to the standard record retention policy. To comply with this policy, every employee would have to review potentially hundreds of e-documents *every day* against a long manual of retention periods, classify the e-documents, and somehow save them with the appropriate retention periods (perhaps by printing them out and putting them in folders). In reality, few if any employees ever look at the manual. This is not because they are bad corporate citizens; it is because to do so would be a significant disruption. The policy is not realistic because it doesn't take into account how people actually do their jobs, the tremendous volumes of information with which they're dealing, or the way electronic information is in fact generated and managed.

Clearly, there will be times when it is appropriate and justified to change how employees manage information to meet compliance goals, but you may want to "pick your spots." Changing employee behavior is difficult, and often requires not only training, but also executive communication and other efforts to change corporate culture. The potential benefit of a compliance initiative should not be viewed in a vacuum; rather, it should be assessed in light of the real world ability to enforce the desired changes, and the costs and impacts of doing so.

### 4. Start with Diligence "Top Down"

It is nearly impossible to establish proactive management policies without first gaining an understanding of the organization's information assets. In attempting to gain this understanding, organizations should take a "top-down" approach based on content classification of electronic content, and not a "bottom-up" approach, focused on paper-based processes. Since the vast majority of information is in electronic form, it makes sense to conduct diligence to inventory the "big buckets" of information in your company, where it resides, and how it can potentially be put into managed repositories. The goal over time is to build content classification into the framework of how information is managed, including as new applications are rolled out.

---

At a high level, such diligence might include interviewing some key IT professionals, reviewing a list of your company's applications (as a window into the large categories of your company's information), and if applicable, leveraging content generated in connection with Sarbanes-Oxley 404 compliance, which often includes systems inventories. You can also do diligence with the lines of business, but rather than trying to understand every document type, stay focused on the most significant workflows that drive the business activities of different functions, as a high-level window into the most important content that is being generated. One other practical consideration is to determine which legacy or historic information will be part of the diligence exercise. Often, a practical approach is to focus on the information generated now and "going forward" and leave the legacy information for another day, unless that legacy data is creating significant pain (e.g., warehouses of tapes).

As discussed further below, such an inventory and diligence is a critical early step in obtaining more proactive control over information for eRetention management purposes, but an added benefit of this inventory is that it is useful for establishing a more proactive eDiscovery process. Rather than waiting for a discovery request to hit and then trying to figure out what information you have and how you should collect and preserve it, such an inventory is valuable collateral that allows an organization to identify "sources" of information up front so that the eDiscovery process can be conducted more defensibly and efficiently.

## 5. Prioritize and Don't "Boil the Ocean"

To achieve perfect and immediate information management compliance is not realistic for most organizations. Therefore, there is a need to prioritize. Such priorities can be identified based on the answers to some straightforward questions, including:

- What information, applications, and workflows are most important to the business and which are causing the most cost, disruption, risk, and pain with respect to eRetention and eDiscovery?
- What information is subject to specific statutory obligations (e.g., Sarbanes, HIPAA, and so on), and how is it currently being managed?
- What information is the most difficult for IT to manage because the storage media is not aligned with the actual service-level requirements (e.g., information is sitting on high-performance storage, but no one ever accesses it; information is on tape and regular access is required, and so on).

Most organizations see some if not all of the following as priorities:

- policy management and discovery of e-mail,<sup>18</sup>
- policy management and discovery of unstructured content (e-documents),
- policy management and discovery of Web content,
- policy management and discovery of other important classes of information based on the company's business or vertical industry and key workflows (e.g., mortgage applications, insurance policies, etc.)

<sup>18</sup> Increasingly, organizations are grappling with the management and retention of other forms of digital messaging, including IM, and voice-over-IP. The technology allows such messages to be managed in essentially the same ways as e-mail (but it is a separate question whether an organization would want to implement that process).

---

## 6. Consider Leveraging Specialized and Focused Consulting Resources

If appropriate, leverage specialized and focused consulting resources that can help identify key priorities, provide unique expertise, and bridge the gap between Legal and IT. Consultants can help build or augment a new workable approach to eRetention management, eDiscovery, or they could focus on the issues associated with a particular application, such as e-mail.

Typically, consultants will spend time assessing the existing state, conducting cross-functional interviews and workshops, issuing surveys, and reviewing policies and processes to identify key gaps and to provide expertise to address areas of significant risk. From a technology perspective, there might be a review of how systems are set up, how information is stored, backed up, and the processes and solutions for archiving.

Alternatively, depending on your internal resources and risk profile, your own employees can conduct some of these activities. If you use consultants, make sure you're getting value for your consulting dollars. Don't pay for elaborate engagements that will do little more than provide a detailed accounting of all the "gaps" without enabling practical solutions. Having a theoretically perfect records retention policy that can't be easily understood and implemented by the employees will likely fail. So use consultants with practical, real-world ideas of how to start to fix the problem.

## 7. Simplify Retention Policies, and Incorporate Reasonable, Repeatable, and Cross-Functional Business Process for Retention, Disposition, and Discovery

### 7.1 The Sedona Guidelines

The Sedona Guidelines: Best Practices & Guidelines for *Managing Information & Records in the Electronic Age*, <http://www.thesedonaconference.org> (2005) is a useful starting point for understanding the building blocks of a defensible eRetention management and eDiscovery program. The Guidelines set forth five principles, which in summary are: an organization should have reasonable policies and procedures for management of information and records, these policies should be realistic and tailored to the organization, it is appropriate to destroy information at the end of its lifecycle absent a legal requirement to the contrary, the procedures for enforcing the policies should be comprehensive, and the policies should incorporate litigation holds. In sum, the Guidelines stand for the proposition that a reasonable and good faith program puts an organization in a position to in fact defensibly destroy information at the end of its lifecycle because the company can establish that such destruction is not illegitimate or reactive, but rather is based on the legitimate business justification that information need not be kept after it no longer has any value.

### 7.2 Simplify Retention Policies

If your organization already has a retention manual, consider simplifying and dramatically reducing the number of retention periods that apply to content in your organization. An organization with a policy manual containing hundreds of different retention periods will have a hard time applying those many retention periods to electronic content. However, a tangible first step to operationalizing these retention and disposition periods is to simplify and reduce the number. For example, an organization that went from 200 retention periods to say 10 could begin to close the gap between what the policy says and how the information is actually managed. IT could begin to think about which of the 10 retention periods should apply to information located in a particular repository, and this could be done in a more automated way, by application, including at the time the information is created. This may mean less granularity and perhaps that some classes of documents would in theory be saved for a longer period of time, but in practice, the granular policies are

---

not being followed, such simplification would be a significant step in the right direction. A key goal is to have a set of policies that you actually enforce, meaning that the information is actually deleted in an assured and defensible way at the end of the chosen retention period<sup>19</sup>.

For many organizations, the reality is that every single e-mail, e-document, and the like, is not going to be reviewed, document-by-document, to designate it as an official record, or not an official record. This is because to do so would be too burdensome on individuals and the process necessary to drive consistency and compliance with this type of aspirational policy is untenable. In the electronic world, the existence or absence of an official designation of an object being a “record” often won’t matter. An e-mail that is definitively not an official record under an organization’s records management policy is nevertheless definitively subject to eDiscovery if it contains relevant information. The real question is—does the content of the e-mail have ongoing business, legal, or referential value? In the electronic world, therefore, increasingly the concept of an individual record has to be examined to determine if that fits with how the information is really created and managed. In some circumstances, a “record” should be defined not as an individual document, but more broadly as “information that has ongoing business or legal value.” A pool of information, rather than an individual document, is managed and retained based on its changing value over time, and then it is disposed.

### 7.3 Integrate the eDiscovery Litigation Hold Process and the eRetention Management Process

The convergence of records and management and legal discovery suggests that the traditional records management program should incorporate litigation holds. This typically takes the form of a generalized policy, which essentially states that, once placed on sufficient notice, there is an affirmative obligation to preserve content reasonably related to the subject matter of a pending case, and that destruction of such content is against company policy.

In support of the litigation hold policy, consider implementing a *repeatable and cross-functional business process* for eDiscovery. Traditionally, attorneys have handled legal discovery case-by-case, with each attorney conducting the discovery his or her own way. There was often no need for the involvement of IT since the attorney could just go to the witnesses and ask them to print out and/or collect the relevant paper documents. Today, the process needs to be repeatable because the risks and costs are too high for it to be ad hoc. The process has to be cross-functional because much of the information may well reside outside the control of individual witnesses in the control of IT, and/or it might not be appropriate to have witnesses make judgments about what to collect.

For many organizations, a repeatable and cross-functional eDiscovery process need not necessarily be elaborate<sup>20</sup>. Important elements of such a process include:

- Establish an IT point person or team to assist with eDiscovery information collections and litigation holds. The IT folks would also help Legal to inventory company sources

<sup>19</sup> Reducing the number of retention periods clearly does not solve all of the issues and complexity, for example that exists with respect to event based retention. (Event-based retention is the notion that a retention period is often dynamic, rather than static, and triggered by an event such as three years after employment termination, the life of the policy holder, and so on.)

Moreover, this is not to suggest that traditional methods of records management are irrelevant or misplaced. At most organizations, there will be a need to apply a mix of strategies, and indeed, if a completely manual and paper-based records management process truly works, then you probably don’t want to “fix” it. That said, with the velocity of information, it doesn’t make sense to try to manage massive amounts of content “document-by-document” with manual processes.

<sup>20</sup> How elaborate the e-discovery process is will depend on a number of factors, including the magnitude and complexity of an organization’s litigation and investigation portfolio, tolerance for risk, technology infrastructure, retention policies, industry regulations, and so on.

---

in advance so that the company is better prepared to respond once a discovery request hits. Once placed on notice, this team can conduct information collections using a repeatable process and an audit trail to maintain chain of custody.

- Train attorneys and paralegals to do two key things when a new case is received: (i) send hold notices to potential witnesses, and save an audit trail of those notices (so that if later questioned, the company can demonstrate its efforts to preserve relevant content), and (ii) notify an IT point person or team, including with respect to applying litigation holds to company information (and potentially information that is managed by third parties, if it is relevant and under the company's "control").
- The lawyers (whether outside attorneys or in-house attorneys, or both) also need to provide guidance to IT about what to collect and hold, based on the attorneys' assessment of what is required in the particular case. For example, the intensity and depth of a data collection and preservation in a small slip-and-fall case would likely be different than the effort required in connection with a large government investigation.
- Good communication between in-house and outside counsel, and a proactive and transparent approach to eDiscovery, including communication with the opposing counsel and if necessary the Court, consistent with what is anticipated under the new Federal Rules.<sup>21</sup>
- As discussed further below, increasingly, companies are turning to automated tools—built into the IT infrastructure—to support the eDiscovery process and allow for more efficient and proactive search, collection, culling, litigation hold, review, and production. Rather than waiting for a subpoena to hit and then calling in a reactive services vendor to manually collect and host content, companies can leverage such tools to support an efficient, proactive, and repeatable process.

By having in place and following a repeatable process, an organization can not only drive efficiency, but also may reduce risk since if there is a discovery dispute or even the inadvertent loss of information, the company can demonstrate that its process is reasonable, consistent, implemented in good faith, and not reactive or intended to hinder the legal process.

## **8. When Assessing Technology Solutions, Build a Simple Business Case Based on IT and Legal ROIs.**

Compliance initiatives can be justified for a variety of reasons, including the inherent value of following the law, and risk reduction. For some compliance efforts, a statute might definitively dictate how long a certain type of information is to be retained and disposed, and precisely how it is to be managed. For example, the Securities and Exchange Commission ("SEC") Regulation 17 C.F.R. 240.17a-4 ("17a-4") sets forth specific retention requirements for broker dealer communications especially on e-mail and IM. In such circumstances, the company must comply. However, many of the obligations relating to eDiscovery and eRetention management are not covered by a definitive statute. Instead, they're driven by the practical question of how long is the information needed for the business, together with less well defined eDiscovery "reasonableness" standards. In this context, as a practical matter, a critical question often is—what is the return on investment for a chosen solution? A business case is often an imperative.

<sup>21</sup> For example, assume that a defendant company collects what it believes to be a relevant universe of information in connection with a case; however, the plaintiff seeks production of a much greater universe of information. Rather than simply saying "no," which might well have been the response in the past, the defendant might consider conducting diligence to quantitatively show how much the next level of discovery would in fact cost, that it is unlikely to yield relevant information, and therefore is not reasonable. Then communicate that information and perhaps a reasonable counter-proposal to the plaintiff. If the plaintiff accepts this offer, then the issue is resolved efficiently. If the plaintiff rejects the offer, then the defendant would likely be in a good position to go before the court and argue that the deeper discovery should not be allowed, or that costs should be shifted to the plaintiffs. See generally, Fed. R. Civ. P. 26(b)(2); *Zubulake v. UBS Warburg*, 216 F.R.D. 280 (S.D.N.Y. 2003) ("*Zubulake III*").

---

Whether on its own or with consultants, organizations can conduct a commonsense analysis of the current state and the potential ROI of a proposed solution. This type of business case might be elaborate or it might be very simple. As a simple example, IT might look at a macro analysis of how much money is being spent on nightly tape backups, and high-performance storage for duplicates and content that is not being accessed. The Legal folks might look at the eDiscovery costs in the top three to five cases from the previous year, which sometimes yield staggering numbers, and compare those to the expected costs if data collections were automated and smaller “universes” were sent to outside counsel. The IT ROI is largely driven by better information management from archive and content management solutions. Those same solutions can drive a corresponding Legal ROI by allowing organizations to automate their data collections, efficiently send a smaller and more relevant “universe” to outside counsel, and achieve litigation hold on selected content rather than saving everything.

## **9. Implement an Enterprise ILM Strategy Broken into Digestible Pieces, Apply a “Big-Buckets” and then “Little-Buckets” Strategy, and Build eDiscovery Efficiencies into the IT Infrastructure**

### **9.1 ILM**

Because the stakes are so high, information management technology decisions need to be made carefully. Traditionally, individual business functions might have chosen their own processes and *point solutions*, but in the electronic world, these decisions may have an impact across an organization. Point solutions might solve a particular group’s short-term issues, but ultimately are inefficient because they’re not integrated, and information ends up stove-piped and difficult to manage.

*Information lifecycle management (ILM)* is a powerful enterprise information technology strategy based on the simple fact that all information is not created equal. Different categories of information require different policy management, and over time, the value of information changes during its lifecycle. An ILM strategy includes integrated services, software, and hardware to allow organizations to policy-manage information, and to align its changing value to the appropriate levels of accessibility, protection, retention, and management.

An ILM strategy allows an organization to utilize the same infrastructure for proactive eRetention management, as well as for eDiscovery, by building tools into the infrastructure that allow for efficient search and content management.

An ILM strategy also allows an organization to *engage in an enterprise strategy, broken into digestible pieces*. As a result, priority information management projects can be implemented over time, but they fit together in a logical, scalable, and integrated way.

### **9.2 Big Buckets and Then Little Buckets**

Within the records management world, there is sometimes a debate about a “big-buckets” versus “little-buckets” strategy for retention management. This debate reflects the reality that setting eRetention management policies across an organization can be complicated, and that in practice there can be an inverse relationship between achieving granular classifications and the impact on the enterprise and its employees. The advocates of a big-bucket approach tend to focus on simplification and low impact on employees. The advocates of a little-bucket approach tend to focus on the importance of the compliance goals, and the need for granular and accurate classification of individual records. The debate, however, is artificial. There is no one-size-fits-all strategy, but there is sometimes a logical progression—“big buckets” *and then* “little buckets.”

For many organizations, a practical strategy is to start by getting information (especially e-mail and unstructured data) into managed archives (“big buckets”) and then over time, apply content management solutions to targeted groups, content types and workflows, where greater granularity of classification (“little buckets”) can logically be implemented, without excessive impact. This approach is discussed further in Section IV (“E-mail as an Example”) below.

### 9.3 Built-in Tools for eDiscovery

With an ILM strategy, the same IT infrastructure that provides for proactive information management can be leveraged to make the eDiscovery process far more efficient. Information that sits in a managed repository, and that is de-duplicated, is far easier to discover than unmanaged information. Federated search allows a company to tap into many types of content across an enterprise and conduct automated and focused data collections, resulting in less content being sent to outside counsel for legal review. When a company has many pending legal matters, trying to manage the eDiscovery process from a warehouse full of tapes quickly leads to a “save-everything” strategy. There is little or no ability to enforce the chosen disposition policy out of fear that something called for in discovery would be inadvertently destroyed. With a combination of archive and content management, litigation hold can be achieved by using a federated search tool to collect relevant materials and lock them down in a matter vault within the content management environment, allowing the company to continue to policy-manage the “source” repositories and avoid the “save-everything” strategy (see figure 1 below).

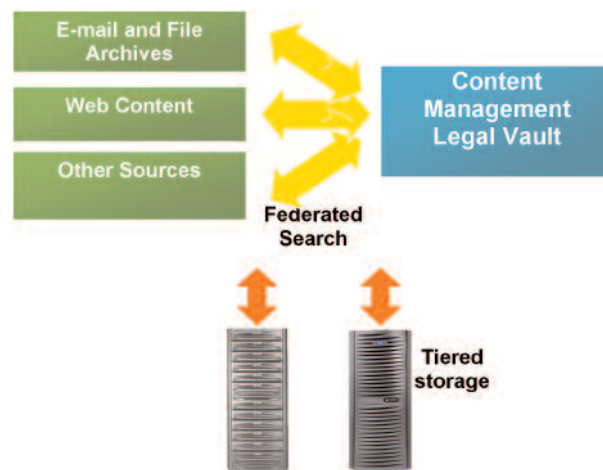


Figure 1. IT Infrastructure Tools to Automate eDiscovery

The creation of a “matter vault” in a content management environment also allows a company to organize data collections by matter (meaning that the information can be kept for the life of the case and then destroyed), set access security controls (on who gets to see the information within a legal matter), create a platform for reviewing and culling down the collected content, maintain the proper metadata and online chain of custody, and efficiently export the responsive information.

## 10. Continuous improvement.

It almost goes without saying that to make the overall program work you need to conduct yearly reviews, audits, training, and cross-functional oversight to refine processes, and apply technology improvements when they make sense. For some organizations, much of this infrastructure may already exist.

---

## Section IV: E-mail as an Example

The following is a practical discussion of e-mail archive and content management solutions, and cross-functional considerations for setting e-mail retention and disposition policies.

### 1. Diligence

If, as is the case for many organizations, e-mail is identified as a priority, a logical next step is to conduct diligence to understand what types of content reside on e-mail, the business value of this content, what regulations, if any, apply to this content, how the e-mail environment is currently managed, what policies are in place and whether they are being enforced. The goal is to understand the current state, the gaps, and the ways that money is being wasted and risk being incurred.

Depending on particular circumstances, industry, geography, and risk profile of an organization, it may make sense to engage outside consulting to do this type of assessment and gap analysis. Other organizations may find it sufficient to conduct their own diligence, including reviewing existing policies, and conducting simple interviews or informal surveys. For example, asking employees across different functions commonsense questions such as: (i) what types of content do you have on e-mail, (ii) how do you manage your e-mail box, (iii) do you have personal e-mail archives, and if so, how often do you actually look at what's on them, and (iv) putting on your company hat, how long do you really need to keep this stuff? Internal IT resources and/or consultants can also assess the current management of the e-mail environment, including the archive and backup policies and supporting processes, and macro analyses of e-mail data management, such as for monthly tape costs.

### 2. Cross-Functional Policy Dialogue

#### 2.1 Classification

Once there is a basic understanding of the current environment and gaps, it makes sense to have a cross-functional dialogue about what policies to apply to e-mail in light of available technology solutions. In essence, this dialogue is a way to define with more precision the problem that the company is trying to solve. Since there are so many different types of content on e-mail, it is helpful to break down the retention policy discussion into component parts:

- How should we classify the e-mail content?
- Once the classification is achieved, what should be the *duration* of the associated retention period(s)?

At a macro level, e-mail environments contain: (i) business content that has some ongoing value to the business, (ii) regulated content that needs to be retained and managed in a certain way based on a statutory requirement, (iii) content that is subject to litigation hold, and (iv) junk that has no ongoing business or legal value. There are of course many subcategories underneath these general headings.

How can the important e-mails be separated from the junk, how granular should the classifications be, and what can the technology do? Although the technology for automated classification is becoming increasingly robust, there is currently no “push-button” solution that automatically classifies all of a corporation's e-mails into a perfect corporate taxonomy. Consequently, decisions have to be made about the level of employee involvement to achieve classification, and the tradeoffs between granularity of classification, consistency, and operational impacts.

---

The most cutting edge technology now allows e-mail and other unstructured content to be archived directly into a content management environment.<sup>22</sup> This unique capability allows e-mails and other content types from across an organization to be managed as part of a unified repository and set of workflows. Once in the content management environment, classification services can be applied “behind the scenes,” lowering the impact on individual employees to make classification decisions. Other approaches to gaining control over e-mail management include “drag-and-drop” and “back-end archiving.” The drag-and-drop approach to archiving puts subfolders with associated retention periods on employee desktops. The employees can then drag-and-drop the e-mails they want to retain into those subfolders, and the e-mails automatically end up in a centrally managed repository, rather than on the user’s hard drive in a personal archive. The “back-end” archiving approach captures e-mails into a central repository and then in an automated way applies a small number of “big-bucket” retention policies, based on business rules. For example, e-mails for all employees in the Finance, HR, and Legal Departments could be saved for a number of years, and all other functions’ e-mails saved for a different period.

The various technology solutions also allow a company to set rules for which e-mails do and do not reach the archive, and for automatic deletions. For example, a drag-and-drop solution can be set up so that only the e-mails moved to the archived subfolders are retained (for a period mapped to the subfolder), but other e-mails that are not moved to a subfolder can be automatically deleted. Similarly, the back-end archiving solutions can be set up so that all e-mails are automatically saved (or journaled) to the archive for the duration of a chosen retention period (thus preventing employees from deleting any e-mails), or business rules can be applied so that categories of e-mails are deleted before they ever reach the archive.<sup>23</sup> Alternatively, users can be given the ability to delete e-mails during a window of time (say 90 days) before the e-mails reach the archive, but all e-mails that are not deleted during this window of time are automatically archived and retained pursuant to the company’s chosen retention policy.

## 2.2 Duration

It is telling that there currently exists no best practice, by industry or otherwise, with respect to how long to keep e-mails. This is because e-mail is not a record type, it is a messaging system with various categories of information on it. In addition, organizations struggle with the massive volumes versus inadvertent destruction dilemma, and the sometimes emotional views about how long e-mails should or should not be kept. At one extreme are those who want to get rid of all e-mails after a very short period of time. At the other are those who think that all e-mails should be kept forever. For most organizations, the appropriate policy lies somewhere in between.

There is sometimes an assumption that “e-mails can never be destroyed.” Leaving e-mails unmanaged often results in e-mails being kept essentially forever because there are so many copies in so many different places and no automated way to enforce a retention policy. Putting e-mails into a centralized, managed archive allows an organization to actually destroy the e-mails at the end of a chosen retention period. As discussed below, whether that retention period is long or short is based on a variety of factors.

<sup>22</sup> EMC Unveils World's First Enterprise-Class Archiving Software Based on a Unified Archiving Platform, [www.emc.com/news/emc\\_releases/showRelease.jsp?id=4279](http://www.emc.com/news/emc_releases/showRelease.jsp?id=4279).

<sup>23</sup> For example, all employee e-mails containing third-party notices (e.g., Bloomberg), or all unsupported attachments (e.g., jpeg pictures) are deleted in bulk.

---

The following are some considerations for those who are inclined to save e-mails forever:

- Is your strategy based on the concern that you might destroy relevant content subject to litigation hold? If so, would that concern be reduced if you had a repeatable process and the tools built into the IT infrastructure to policy manage and automate the collection and hold of your e-mails?
- Have you considered separating your archive and disaster recovery backup strategies?
- Are you looking at the e-mail environment monolithically, or have you done some diligence to understand the “big-bucket” categories of information that you have on e-mail, so you can consider mapping logical policies?
- If a particular content type is driving your desire to save everything, have you considered whether it would be more efficient or appropriate to move that content off of e-mail?<sup>24</sup>
- Does it make sense to put in place a going-forward solution with a long, but not indefinite, retention period as a way to get started, and to position your company for defensible disposition of e-mails in the future?

The following are some considerations for individuals who are inclined to have a very short e-mail retention period (in essence to use e-mail only as a short-term messaging environment):

- Are you just paying lip service to a short e-mail retention policy, but in reality allowing various workarounds (such as personal archives) so that much of the e-mail content will not in fact be destroyed within the “official” policy period?
- Do employees currently put lots of important business information on e-mail and save e-mails for a very long time? If so, how will employees respond, and what lengths are you willing to go to change their behavior? Will “good” employees resist your policy, and if so, what does that say about its viability?
- If you are considering e-mail archiving, have you embraced the reality that you can actually enforce a retention period on e-mails in the central repository (so for example, if you set a 90-day policy, the e-mails could really be destroyed in 90 days, with little or no chance to recover them, and do you really want that)? Are you, and others in your function, willing to live by the same policy that is being imposed on others in the company?
- Do you have regulatory content on e-mail, such as Sarbanes or HIPAA content? If so, would you be violating statutory obligations by enforcing a short e-mail retention period?
- If you impose a short retention period, what will employees do with all of the attachments? Do you have a complementary content management solution that allows employees to easily save attachments so that they can be policy-managed and retrieved in an automated way?
- Is your eDiscovery litigation hold process efficient enough to ensure that relevant e-mails are preserved before they get destroyed pursuant to a short retention policy?
- Is your concern the potential “smoking gun” e-mail? If so, does this justify getting rid of all e-mails after a short time, even those with valuable content or that could provide context that would be helpful in future litigation? Are you better served by training employees on proper use of e-mail (recognizing that some employees will still say dumb things on e-mail, but to do so would be against an enforced company policy)?

<sup>24</sup> For example, an organization might make a policy decision, supported by process, to remove all HIPAA content from e-mail so that they don't keep all e-mails based on a statutory requirement that applies only to a subset of the e-mail environment. Such a decision would have to be assessed realistically, included with respect to the impact on employees and the ability to implement it, but for some, those impacts would be justified to get to a shorter retention period.

---

In sum, by having a cross-functional discussion about retention periods, breaking the discussion down into its component parts, and embracing reality, you can define in a realistic way what problem you are really trying to solve, and you can establish and enforce a realistic retention policy.<sup>25</sup>

### 3.1 Big Buckets and Then Little Buckets.

The benefits of implementing an e-mail archive are compelling. They include: (i) cost savings from the consolidation of a fast growing and unmanaged environment; (ii) de-duplication and tiered storage (automatically mapping storage capabilities with the accessibility and other service-level needs based on the lifecycle of the content); (iii) automation of data collections across the e-mail repository for eDiscovery purposes; and (iv) the ability to enforce a retention policy for e-mail.

In light of these benefits, for many organizations, the fact that the chosen e-mail retention policies will not be perfect is not a sufficient reason to leave e-mail unmanaged. Some e-mails will be saved that you don't want to be saved (e.g., idle chatter), some will be kept longer than you want (e.g., routine business communications), and still others might have content that should be kept even longer than the chosen retention period (e.g., contracts, regulatory content, and trade secrets). Embrace reality. If in actual practice there has been no enforcement of an e-mail retention policy (for example because of personal archives or tapes), then even a seemingly "long" retention period of years and years may be a step in the right direction. As a practical matter, a longer initial e-mail retention period is easier to implement because the impact on employees is lower. This does not mean that an organization must save e-mails for excessively long retention periods, but this is a useful and realistic consideration.<sup>26</sup>

This can be viewed as a journey. Once you have information in managed repositories and a reasonable "big-bucket" policy has been applied, for example, three years for a particular function, this creates a three-year window in which a company can refine the policies and implement more granularity. This can be done by applying complementary *content management solutions* and associated services where they make sense, including by looking at who within the company is amenable to such solutions, and which content and workflows are priorities.

### 3.2 Amenable Functions, Priority Content, and Logical Workflows

Certain functions within an organization tend to be process-oriented (e.g., procurement, finance, quality, and so on) and therefore more amenable to embracing records management. Perhaps all employees across an organization will not be consistent about classifying their e-mails or e-documents, but for example, a procurement group might be given "drag-and-drop" functionality that would allow them to take a final contract out of e-mail and put it into a content management environment, where services can be applied to build a detailed file plan that allows for granular classification. Thus, this group would get helpful content management tools, and from a compliance perspective, correct and granular policies would be applied to a priority record type.

<sup>25</sup> In addition to the retention period, there are often a number of other policy questions including those relating to security and privacy. These policies may be driven by either statutory requirements or by business rules. Although a detailed discussion of security and privacy issues are outside the scope of this white paper, it should be noted that the same methodology can be applied to address these issues. In general, you need to conduct diligence, prioritize, classify content, and apply corresponding business rules as part of an enterprise strategy. For example, once personal financial or medical content is identified, the proper policies relating to access control, audit, data immutability, assured destruction, encryption, and so on can be applied by leveraging a flexible ILM infrastructure.

<sup>26</sup> It is sometimes uncomfortable from a records management perspective to come to terms with the fact that an e-mail policy is being set for a duration, but at the same time there is content in e-mail that is subject to a different retention period pursuant to the existing manual of retention periods. First, this underscores why policy simplification is a worthwhile goal. Second, it highlights risk management and the need to embrace reality. If in actual practice thousands of individual employees are saving e-mail content for thousands of individually determined retention periods that do not line up with the manual, then going to a small number of enforced e-mail retention periods is likely an important step toward far better compliance.

---

As in the example above, certain content (contracts, regulatory content, trade secrets, critical business records, etc.) is sufficiently important that granular content management is needed. This may require some manual process and the involvement of the employees who create this content, as well as the application of automated classification tools. One of the critical benefits of an ILM strategy is that additional services and automation can be applied over time to information that resides in the content management repository.<sup>27</sup>

Finally, organizations tend to have certain key workflows, which are good candidates for content management automation because it is more efficient to automate the process. For example, assume a team of engineers uses e-mail to communicate about development of software code, resulting in the company saving those engineering e-mails forever out of a fear of destroying trade secrets. A content management and collaboration solution would require those engineers to “check-in” and “checkout” software code, apply version controls, and implement an automated workflow so that team leaders can review and lock down final code versions online. The engineers may be willing to implement this solution because of the potential business efficiencies of automating the apply version controls, and implement an automated workflow so that team leaders can review and lock down final code versions online. The engineers may be willing to implement this solution because of the potential business efficiencies of automating the process and workflow, but there are additional compliance benefits. All the documentation generated by this workflow would end up in a content management environment where it could be classified and appropriate policies applied. Moreover, as the trade secrets are removed from the “big-bucket” e-mail environment, the company may over time be able to reduce the e-mail retention period for these engineers.

## Section V: Conclusion

The challenges and complexities of eRetention management and eDiscovery compliance can seem massive, but with cross-functional communication, prioritization, risk management, repeatable process, and an enterprise ILM strategy broken into digestible pieces, organizations can tackle the core issue of proactive information management, and reduce costs and risks.

<sup>27</sup> Content can be managed within a content management archive, or it can potentially be managed “virtually” by having policies applied even to content that sits outside of managed repositories.

---

*EMC Corporation (NYSE: EMC), a Fortune 500 company, is uniquely positioned to help companies address the eRetention and eDiscovery compliance challenges at their core—with proactive information management, together with the full range of focused and domain-specific consulting and professional services. EMC is the world leader in products, services, and solutions for information management that help organizations extract the maximum value from their information, at the lowest total cost, across every point in the information lifecycle. EMC helps enterprises of all sizes manage their growing volumes of information—from creation to disposal—according to its changing value to the business through information lifecycle management (ILM) strategies. EMC information infrastructure solutions are at the heart of this mission, helping organizations manage, use, protect, and share their information assets more efficiently, cost-effectively, and in a compliant manner. Information about EMC's products and services can be found at [www.EMC.com](http://www.EMC.com).*



**EMC Corporation**  
Hopkinton  
Massachusetts  
01748-9103  
1-508-435-1000  
In North America 1-866-464-7381

EMC<sup>2</sup>, EMC, and where information lives are registered trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.

© Copyright 2006 EMC Corporation.  
All rights reserved. Published in the USA. 4/06

Data Sheet  
H2153