5-2016

# Email Archive Retention Using Enterprise Content Management

Harini Muppidi

# Email Archive Retention Using Enterprise Content Management

by

Harini Muppidi

A Starred Paper

Submitted to the Graduate Faculty of

St Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Engineering Management

May, 2016

Starred Paper Committee:
Ben Baliga, Chairperson
Hiral Shah
Balasubramanian Kasi

## Abstract

The company is one of the leading nation's largest property causality companies, which has more than 30,000 employees. By implementing an email, archive management system the company categorized the value of emails and organized them to increase the productivity, retention period and decrease the chances of litigation. The company had chosen IBM FileNet out of three vendors and accomplished the goals of the project.

**Acknowledgements**

This project completion is mere impossible without the assistance and valuable guidance from many individuals.

I am extremely thankful to Dr. Ben Baliga, Professor and Graduate Professor, for Engineering Management Program at St Could State University. His support, guidance, and assistance assisted me to complete this project successfully.

I would take this opportunity to thank Dr. Hiral Shah, Associate Professor for Engineering Management Program, at St. Cloud State University for her support and guidance.

Also I would like to thank Prof. Balsy Kasi for serving on the committee and support throughout the project.

## Table of Contents

Chapter Page

## List of Tables

## List of Figures

# Chapter I: Introduction

## Introduction

Email is without question a revolutionary communication tool that transformed the way a business is done. Poorly managed email represents significant corporate risk and potential liability. Organizations must secure and retain email records or face fines for the illegal destruction of email and sanctions related to non-compliance. The challenges associated with mismanaged email are multifold and complex, least of which is the substantial IT cost and complexity associated with maintaining the mail environment, and the frustration faced by end users who are overwhelmed by the volume of email they deal with on a daily basis.

Enterprise Content Management (ECM) is the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. Using ECM, the E-mail archive management initiative moves the company away from a legacy e-mail archive solution to a strategic information governance solution. The project was implemented at nation's one of the largest property causality companies, which has more than 30,000 employees.

## Problem Statement

Email servers were never designed to act as repositories for huge quantities of emails and move control of this information away from the organization. The more information generated, and in turn saved, the greater the cost and risk that will incur throughout the litigation and other discovery-driven processes.

The company was compelled to resolve two fundamental email challenges,

- Automate and safeguard the disposal of unimportant and low value emails.
- Retain business value and intended record email according to the company policies.

**Nature and Significance of the Problem**

The vast quantities of emails held in inboxes, sent folders, and deleted items folders put the organization at risk and adversely impact the performance of email servers in the organization. There is high risk of catastrophic loss of information and increased exposure to litigation, investigation and audit. Email management systems centrally capture emails created and received by employees. Using a classification scheme to manage this content, retention periods and access controls can be applied to manage emails. Email management involves the systematic control of the quality and quantity of electronic messages that are sent from within, and received by, an organization. The implementation of this project benefited the company by:

- Improving defensible management and disposition of email.
- Classification of emails for records retention purpose.
- Provides reliable and compliant personal email archive.
- Enables the ability to save e-mail based on its content and value rather than saving all emails indefinitely.

**Objective of the Project**

The main objective of the project was to provide the end user with an email archival solution that allows them to do their work in an efficient and effective manner.

The other objectives of the project include:

- Establishing the requirements for email classification, legal holds and automated disposition, and
- Implement a solution to support classification, legal holds and disposition that will align with the companies record retention policies and supporting current business processes and systems.

**Project Questions/Hypotheses**

- How will the current solution make e-mail records compliant?
- How does the project ensure personal identifiable information is secure?
- How long would the emails be retained?
- Was the retention schedule acceptable by end user?
- Can we have another simplified solution?

**Assumptions**

The assumptions that were made prior to the initiation of project were,

- There will be 1 Pilot that lasted for 4 weeks
  - Pre pilot-Core Project team
  - Pilot 1: Corporate systems, Corporate services and Underwriting
- Rollouts will occur by business unit

- Checklists will be used to certify areas for their rollouts

**Limitations of the Project**

- Availability of knowledgeable Records Management resources
- Only certain team members will provide technical support throughout the rollout
- Communications will be managed by Enterprise communications

**Delimitations**

The project was focused on retaining the emails to only certain period of time, in this case being 2 years.

**Summary**

This chapter gives the brief introduction about the project. This chapter is meant to introduce the aspects of the project like scope, objective, nature and significance of the problem. This is the preliminary analysis to be done before the initiation of the project.

# Chapter II: Background and Review of Literature

## Introduction

This chapter is focused on reviewing the literature related to the problem. This chapter introduces to the reader about the search strategies and the background related to enterprise content management as well as email archiving. It further focuses on the literature related to methodology.

## Background Related to the Problem

The company is nation's one of the largest property casualty companies. It has more than 30,000 employees and 13,000 independent agents and multiple market segments across the personal, business, financial and international insurance groups. It is spread worldwide with different places of operation (retrieved from source website of the company).

The email archive project was a joint effort between company's legal and record management teams that was introducing an Email-based Record Management archive solution. The solution consists of a FileNet plug-in that snaps into each users Outlook application. This plug-in was supported by a set of backend FileNet servers that maintain the email record, allow the record to viewed, retrieved and eventually be disposed based on its specific retention definition.

Before the implementation of the project, much of the unstructured information at the company was managed at the individual user level. The individual determines if the information should be stored within the email system, on portable media, personal or shared folder, or printed and stored in a file room or cabinet. Email

continues to grow exponentially and there were no standardized space restrictions in place. Users have the ability to create. PST files and to decide where they are stored. All email deleted by individuals were retained indefinitely in the corporate archive. There was no effective system in place to assist users with content -based management, retention and deletion of email.

In the event of electronic legal discovery, IT was tasked with identifying and collecting information from the myriad of storage locations. Since the collected information was matter specific, the same information was being collected multiple times. Because emails were not classified (categorized), preservation and collection were done at the mailbox level, which was far more than necessary, resulting in unnecessary expense.

The company has previous Records Management policies standards, procedures, training and a Records Retention Schedule to govern the management of information, but the platform does not allow for classification of email according to Records Management policies. The email archiving platform solution at the beginning of the project was an older version of the EMC solution. The architecture was not scalable to the volumes the company has reached of 130+ terabytes. The volume of the companies retained email continues to grow at about 30-40% per year and all emails were saved indefinitely. There was a lack of stability due to index corruption as well as other factors. Recovery of large corrupted indexes took several days or longer, increasing the risk of failing to meet e-Discovery SLA's. In order to keep indexes manageable, the company has had to deploy an extensive infrastructure.

The company continued to add to this infrastructure to accommodate growth. Search capabilities were weak, as the archive function does not support complex searches, keying / tagging data to support searches or searches that complete in a reasonable time frame (some take days / weeks). This often resulted in missed SLAs and large data sets being returned, which in turn requires external legal services to narrow search results to relevant data, As the company moved forward, the previous solution became more difficult to maintain and continued to miss SLA's for availability.

In assessing the business problems with the existing email archive, the company' Core Team identified the severity for 11 major challenges. Business problem with Email Archive before implementation of project are listed below in order prioritized from critical to low severity

**Critical severity**.

1. The archive size, stability and ability to search:
    a. Cannot execute timely complex searches due to archive data volume and indexing issues.
    b. Not all emails were captured due to insufficient controls between archive and email system.

**High severity**.

2. EMC vendor primary support for previous legato archive solution ends soon:

   a. EMC legato primary support was to have been ended soon unless extended further. Customers interested with uninterrupted support has to be upgraded to then current release or purchase extended support.
   b. At the time of project initiation, hardware/software spend was approximately $700,000 annually.

3. Emails retained indefinitely resulted in additional storage costs:
   a. Over 1 billion emails stored in EMC, with additional 1 million saved each business day, utilizing about 130+ terabytes of data with expected growth of 10 terabytes annually.
   b. User archives email utilizes 20 terabytes across 160,000 PST files.
4. Poor search capability with previous archive results in larger data sets being returned than needed:
   a. Increased internal staff expense and outside vendor service fees increase given size of datasets from our searches.
   b. E-discovery spend was approximately $10MM per year, potential exists to save $2-$3MM annually on outside service fees.

**Medium severity**.

5. Existing solution does not allow identifying emails subject to legal holds:
   a. All emails had to be held indefinitely since preservation obligations were unknown.

6. Emails retained created unneeded legal exposure:
    a. Disposing of emails per records retention policies would lessen enterprise to litigation exposure.
    b. $5-$10 million potential fines and penalties, including damage to the company's reputation could be realized with the current practice of saving all email indefinitely.
7. Emails retained indefinitely raises costs for retrieving data:
    a. Having emails stored in both EMC Legato and in PST files, and also having larger amounts of data to search than needed increased costs.
8. Inability to manage email by content in order to comply with retention policies and legal holds:
    a. Archived email should be retained and disposed as per the company retention policy..
    b. Reduce organizational risk due to the enhanced ability to locate and retrieve archived email.
9. Existing solution does not provide a centralized way to store or retrieve IM or voice records:
    a. E-discovery requirements now extend beyond emails to instant messaging, live meeting recordings, and voice mails.
    b. Apply consistent information governance solutions of IM, live meeting recordings, and other collaborative information.

10. Emails retained indefinitely increases risk that will not meet e-discovery:
    a. E-discovery requests are a time sensitive activity and the inability to meet discovery order timelines has resulted in unneeded legal fees.
11. PST usage creates customer service issues:
    a. Eliminate PST files and provide a more effective personal archiving process for email end-users.

**Literature Related to the Problem**

A typical information worker who sits at a computer all day turns to his e-mail program more than 50 times and uses instant messaging 77 times. The fractured attention comes at a cost. In the United States, more than $650 billion a year in productivity is lost because of unnecessary interruptions (retrieved from www.nytimes.com). It is highly important to provide a well-defined system that can increase the productivity and provide litigation readiness. ECM provides the company with the ability to manage email on a consistent basis by reducing the volume by deleting obsolete and redundant emails on an ongoing basis.

**Enterprise content management**. Enterprise Content Management (ECM) is the management of information in all its forms across an organization. This aims to capture, preserve and deliver information as a corporate asset in a consistent, natural and re-usable way, so that an organization can sustain, enhance and tune its knowledge investment. Apart from this management, ECM refers to the related strategies, methods and tools. ECM tools and strategies allow the management of an

organization's unstructured information, wherever and whenever this exists (Kidda, Grego, Caplinger, & Houberg, 2011).

The enterprise perspective describes all the functions of distribution, application, publication, and acquisition, capture and access in a uniform and pervasive nature without boundaries. It defines where and how ECM takes effect. The content describes all the rich components, information, data (structured or unstructured), records, rules, structures, topics and templates. It defines what makes up ECM. The management discipline brings together facets of communication, processes, workflows, collaboration, interaction and exchange with a plethora of stakeholders. It describes who is involved in ECM, and why and when they interact (Cameron, 2011).

In the future ECM aims to ensure that repositories of the internet and organizations become federated, consistently searchable, shareable, verifiable and persistent sources coalesce ideas into actionable, valuable knowledge through collaboration protect organizations' ideas whilst sharing and fostering those appropriate for development in the public domain. The Internet has created both cohesion and fragmentation. It has made the globe smaller, breaking down old organizational walls by using a common protocol. In the new world there are no boundaries of country, race, class, gender, religion or government.

Figure 1 describes the structured and unstructured data in an organization.

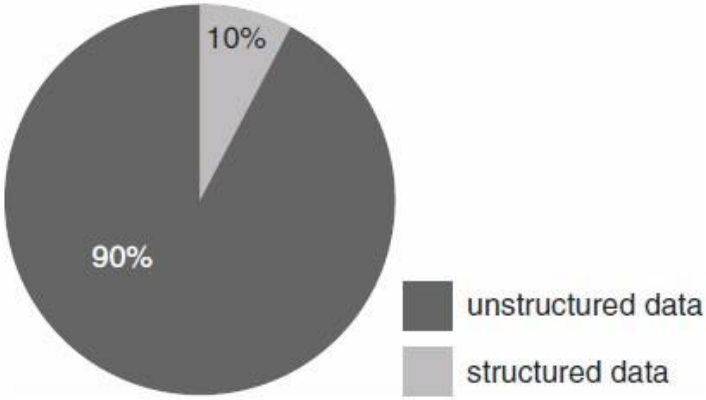


Figure 1: Structured and Unstructured Data

**Email management**. Interest in email management has increased due to major corporate collapses, some involving allegations of fraud, malpractice and exposure through email communication. Casual and business related email conversations link business and personal worlds, so users often have difficulty distinguishing what is an official record. Depending on the habits of the user, emails may be deleted immediately after reading, or retained for years. It is not uncommon to find users with many thousands of mails kept in bloated 'Inboxes'. There are users for whom filing or deletion are seldom if ever performed. A rule-based email management solution significantly simplifies the management of corporate email, and is totally transparent to the end user.

Email Manager provides a fully automated and centralized email capture process. Messages and their attachments matching specific business rules are automatically indexed and archived in the FileNet repository. Important corporate emails can be retrieved when they are needed to address legal discovery requests, collaborative activities or regulatory requirements. The ability to automatically capture

messages is a cornerstone to reduced cost of ownership. Automatic capture reduces the chance of error and the workload on end users. Automation does not stop with the capture process, but extends through the entire life of the message. Messages are kept for the required period, destroyed when they should be and discoverable when it counts (Brogan & Vreugdenburg, 2008).

**Email archiving**. Email archiving is a system approach to saving and protecting the data contained in email messages and IMs so it can be accessed quickly and reliably at a later date. Protection includes accidental loss and deliberate attempts to alter or delete emails. Implementations vary from simple software running on existing systems to specific dedicated hardware. Applications can be built on top of an email archive to address email legal discovery, internal and external email audit support, and many other systems.

A White Paper by Message Labs (2009) drew attention to the fact that more than three quarters of an organization's business-critical data is in email form. The main body of evidence that put Arthur Anderson out of business in 2002, after 89 years, came from email. Citigroup paid $400 million in fines after Elliot Spitzer subpoenaed emails written by stock analyst Jack Grubman. The stock price of the insurance broker Marsh and McLennan dropped by 50% in 2004 after emails showed evidence of kickbacks. In 2005, a federal judge recommended entry of a default judgment against PriceWaterhouseCoopers for deleting emails relevant to a $139M shareholder suit (Floyer, 2009). This explains all how much email archiving is important not only for effective workflow but also for establishing legal holds. The

email archiving team must assess the actual scope of the risk, as measured by potential losses from three main areas–fines from noncompliance, punitive damages, and lack of accountability leading to potential extensive criminal witch hunts. The scope of these risk elements will vary by industry, geography and the need to sustain a history of email activity.

**Literature Related to the Methodology**

The work stream adhered to the SDLC framework with certain exceptions that was handled via executive steering committees. The other major framework was agile methodology.

**SDLC.** SDLC is the acronym for Software Development Life Cycle. SDLC is a process followed for a software project, within a software organization. It consists of a detailed plan describing how to develop, maintain, replace and alter or enhance specific software. The life cycle defines a methodology for improving the quality of software and the overall development process. A typical SDLC consists of the following stages

a. *Program and project planning:* Program and project planning is important as it describes the necessary planning for software and system efforts during software design/development life cycles. The definitions of systems design, software requirements and design, configuration control, systems and software integration, subcontractor involvement, deliveries, and product quality evaluations are critical to effective planning efforts. The initiation of planning starts at the proposal phase with the customer. The result of defined software design/development plans, processes,

procedures, subcontractor support, and effective software tools provides estimations for cost and schedules to be available for teams that are impacted from the start of the proposal phase to delivery of the work products to the customer.

Before a program can require a plan, program objectives are defined and technical and management disciplines are identified. This information defines a reasonable estimate or cause for cost evaluations, risk management assessments, manageable schedules and progress reports. The main reason that software projects are planned and controlled is to eliminate any confusion that could occur. The teams that are expected to provide work products struggle if projects are not planned, and control is not even an option. Studies showed that when schedule, cost, and quality objectives are not a top priority, the project is not successful. Communication planning principals define goals and objectives during the course of program and project planning. The planning aspects require a set of managers to understand not only their position but also the technical practices that support systems and software engineering and to define the course that lies ahead. There are many planning ideas and decisions by managers that are not accepted by team members due to the complexity of change. As the part of project planning the roles and responsibilities were assigned in the project.

Table 1 clearly defines about the roles and responsibilities.

Table 1: Roles and Responsibilities

| S. No | Role | Responsibilities |
|---|---|---|
| 1 | Schedule Management-Primary AD enrollment-Backup | Develops and manages the wave deployments<br>Adjusts the schedule as needed for area changes |
| 2 | Legal Sponsor | Executive and mid-level communications<br>Project advocacy<br>High -Level schedule negotiations with contacts<br>Legal Advisor |
| 3 | Records Management Sponsor | Executive and mid-level communications<br>Project advocacy<br>Project oversight |
| 4 | Enterprise communications-Project communications lead | Review communications plan materials<br>Sends out all projects related communications<br>Adjust communications as changes are needed |
| 5 | IT Education | Creates Training materials |
| 6 | Template/Folder Management | Migrate all architecture components into production<br>Provide on-going technical support for templates<br>Provide Cognos reporting support<br>Identify and assist with upgrade activities<br>Technical Issue resolution |
| 7 | Legal IT support | Provide rollout support<br>Managing subscriptions rule definitions<br>Technical Issue resolution |
| 8 | Cognos Reporting Support | Migrate all architecture components into production<br>Provide Cognos reporting support |
| 9 | Messaging support | Coordinate desktop deployment<br>Coordinate altering inbox retention with rollout waves |

*b. Defining requirements:* Identifying and defining software requirements begin with reviewing the functional or performance requirements developed to identify the constraints on software. The system requirement that is allocated to software evaluations determines accuracy, completeness, and applicability of the requirements for work products.

Figure 2 represents the complete process of SDLC.

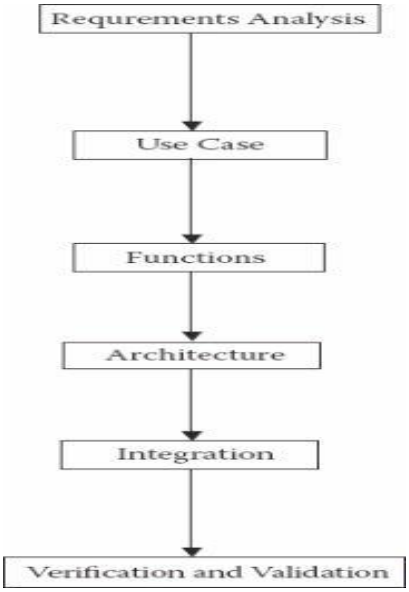


Figure 2: SDLC Process

As the part of requirement gathering, these are few requirements that were approved,

i. All systems comprising the solution and their components must be integrated. The system must integrate within the company's architecture.
ii. Must be able to manage information and support access and management of companies existing email environments.
iii. must be readily available to users and custodians. Must provide the ability to import personal email archives into a central repository.
iv. Solution must have minimal version updates.
v. Data.

*c. Designing the product architecture:* Software design is a consistent approach and method for the development of software requirements in defined designs of a work product. The software architecture definition provides a framework

for the creation of the product design and at times can provide constrictions. The software design definition implements details about a software product's architecture, components, and interfaces. Software designers use element traceability of the design and the software requirements. The traceability data and software design definitions are documented according to program and project plans, ideas, processes, and procedures and applicable internal work instructions.

Figure 3 shows the architecture diagram of the future record management.

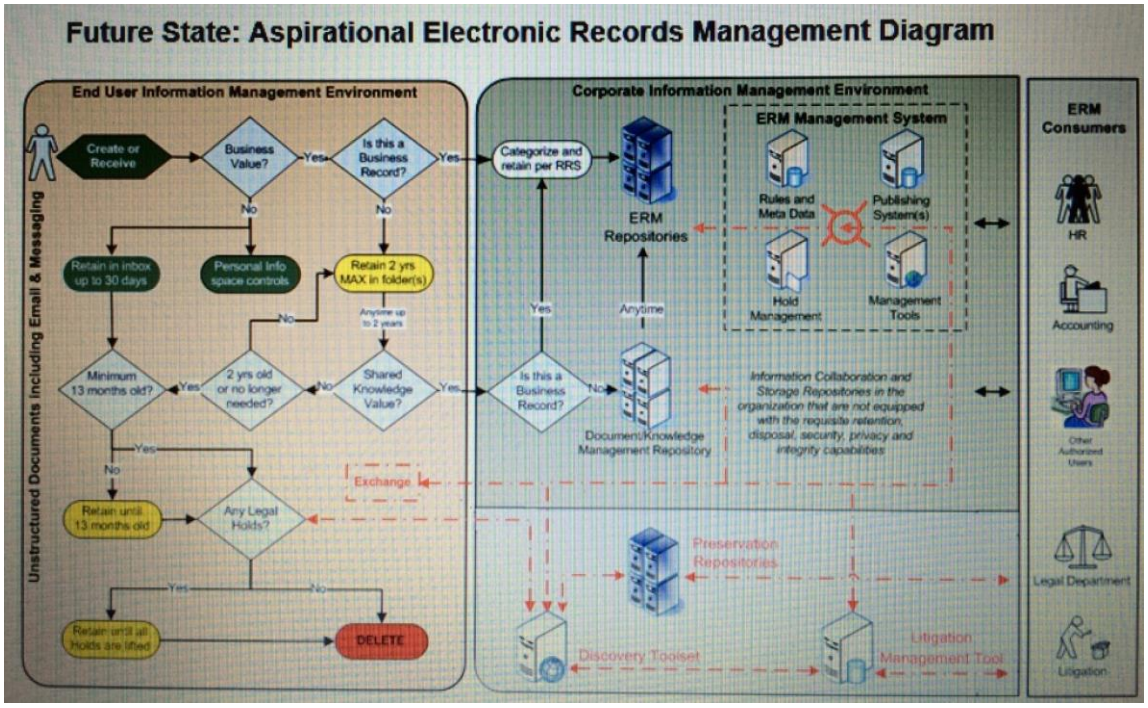


Figure 3: Future State: Aspirational Electronic Records Management

*d. Building or Developing the Product:* In this stage of SDLC the actual development starts and the product is built. The programming code is generated as per DDS during this stage. If the design is performed in a detailed and organized manner, code generation can be accomplished without much difficulty.

*e. Testing the Product:* This stage is usually a subset of all the stages as in the modern SDLC models; the testing activities are mostly involved in all the stages of SDLC. In this stage the products defects are reported.

*f. Deployment in the Market and Maintenance:* Once the product is tested and ready to be deployed it is released formally in the appropriate market. Sometime product deployment happens in stages as per the organizations. The product may first be released in a limited segment and tested in the real business environment (UAT- User acceptance testing).

There are different SDLC models and the one that this project will be using is.

**Agile methodology**. Agile SDLC model is a combination of iterative and incremental process models with focus on process adaptability and customer satisfaction by rapid delivery of working software product. In Agile the tasks are divided to time boxes to deliver specific features for a release. The most popular methods include Rational Unified Process (RUP), Scrum.

The Agile manifesto principles include,

- Self-development and motivation are important.
- Demo working software is considered instead of just depending on documentation.
- Continuous customer interaction is very important to get proper product requirements.
- Quick responses to change and continuous development.

## Summary

The main purpose of the project was to reduce the exposure to litigation, investigation and audit. An equally important purpose was to provide reliable and compliant personal email archive and classification emails for records. Most of the published literature agrees to the fact that email retention is extremely important for an organizations well-being. Many authors have done intense research for understanding the purpose of email management and the necessity of email archiving. Brief rationale of the methodology is also discussed in this chapter. Further, detailed explanation can be read in the following chapter.

# Chapter III: Methodology

## Introduction

This chapter introduces a detailed understanding about the design of study and data collection process. Timeline and budget involved in the project will be discussed.

## Design of the Study

The approach was both qualitative and quantitative. The study design for this project is divided into four phases.

a. *Phase One: LOB (line of business) Transition Planning*

- Each LOB appointed a LOB contact to coordinate email management implementation, communication and training.
- An on-boarding questionnaire was submitted to each LOB contact to collect information about the readiness of their LOB to implement. From this information, the following issues were addressed.
- Analysis of business impact and readiness for email management implementation.
- Special process development, as required (how users will work with new system).
- Agreement on an implementation Plan and Schedule.

b. *Phase Two: Start Implementation Wave Rollout*

- Begin communications and training program.
- On boarding communications.

- Deployment communications.
- Variety of education sessions on training people to use system correctly.
- Release the following email management changes all at once:
  - Change retention from 30 to 90 days- Zone 1
  - Release the assign policy button- Zone 2
  - Push out the SAVE AS A RECORD to all desktop via machine names- Zone 3

c. *Phase Three: Go Live*

- Support Network in place
- LOB contacts and early adopters
- Help Desk- scripts and escalation process

d. *Phase Four: Post Go-Live*

- Content navigator training will be added to the new-hire orientation and new employee, records management training materials will be updated
- Reports will be utilized to monitor appropriate employee usage.

➢ The risks associated with the project were:

- User acceptance of the solution
- Unexpected technical issues
- High level of exception requests

**On-boarding checklist**. As a process of project, an on-boarding checklist was made and was utilized to help Deployment Coordinators and functional areas/

departments for their implementation. The on-boarding checklist had these types of checkpoint activities and the associated timelines that they were supposed to be completed prior to each areas cutover date:

a. Three weeks prior to implementation:
    - Validate AD groups and users associated to each AD group
    - Validate templates, metadata, etc.

b. Two weeks prior to implementation:
    - Validate that the area knows how to escalate issues, questions, etc.
    - Validate that the area has their support materials

*Other on-boarding activities:* A number of other areas either were involved in the support of each wave deployment or have a need to know about each wave deployment. These types of on-boarding activities was conducted and their statuses tracked for each wave deployment:

- Notify the technical consultants about exactly who will be in each wave.
- Co-ordinate with desktop team on the release of the plug-in for each wave.
- Co-ordinate with the FileNet team on the release and complexity or possible amount of new templates that would be needed to maintain timelines.
- Coordinate with a messaging team on the list of users and ID's and date of implementation. This group will be responsible for PST freezes, enabling the Assign Policy button and altering the retention of the inbox from 30-90 days.

*Timeline:* During the designing process activity timelines were mentioned. There were two different timelines Pilot and Full Deployment

Pilot:

- Pilot Timeline: This would need 30-60 days

Full Deployment:

- Rollout Timeline: This would require about one year

**Data Collection**

*Proposed 'Value Management' Approach:* Over the last 20 years, business communication has been revolutionized with email becoming an indispensable tool. Email is a critical application for any business transactions and internal operations. Consequently, email messages play a significant role as evidence in legal proceedings and are subject to costly and time-consuming legal discovery endeavors. The company was required to preserve email for long periods, demonstrate compliance with external regulations, adhere to internal policies, and prepare for possible legal discovery requests. However, the volume of information creates email storage management and system performance problems. The company intended to manage email on a centralized and consistent basis, reducing legal risks.

The company also consistently managed official company records distributed in a plethora of locations. Relevant information was declared as business records by applying and enforcing record management policies and procedures. The company recognizes effective email management, records management, and e-Discovery

involve integrating email, existing and new technologies, business content, transactions, policy and processes. The proposed process required all email to be retained in central repositories, managed at the company corporate level to allow management and control by a centralized record management application. Users would remain responsible to declare email as a record and classify them according to guidelines and categories established by the company Records Management.

The company provided automated tools to assist the user in record declaration and classification. To reduce user burden and maximize user compliance, the new email management solution will focus on identifying and encouraging careful management of the minimal number of emails with high business, legal and/or regulatory value. The email volume was categorized into three values. Below is the representation of various categories of emails

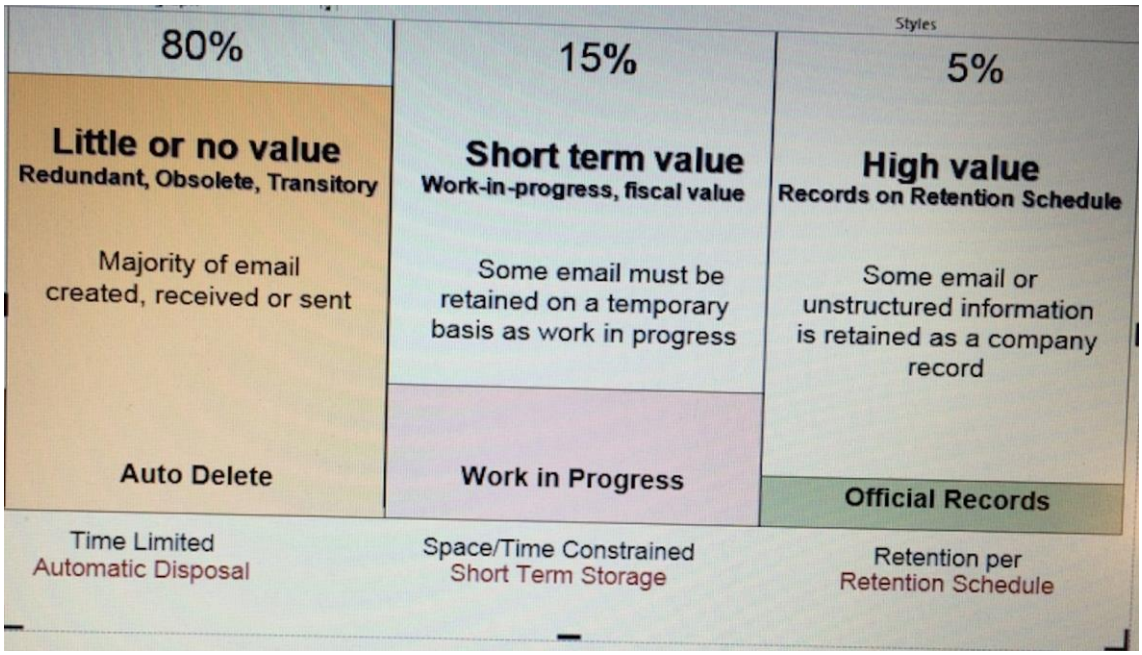


Figure 4: Value Categories of Email

In the event of a legal action requiring a legal/litigation hold and/or e-discovery of company information. The company's legal would be able to search the central repositories and other information stores. This information would then be locked, so users cannot delete it or as scheduled in accordance with the Records Retention Schedule. In the event it is impossible to lock the information in place, the information would be collected.

The key aspects of the 'Value Management' approach are as follows:

- *Auto Delete:* Email that is not moved out of the inbox by users in 30 days will be considered to be "unneeded" information and will be disposed of regularly in the normal course of business. However, all email will remain available for corporate purposes, whether deleted or not, for a minimum of 15 months.
- *Work in Progress.* Users will have reasonable space for working messages and attachments. Storage limits and/or time constraints on these "work-in-process" areas motivate users to actively manage emails, through deletion or declaration and storage as company records. Users will also have some flexibility on how to organize their "work-in-progress" information.
- *Official Records.* Records are "declared" by users and stored in a way that the organized in record categories. These categories directly relate to business and regulatory requirements that provide for the appropriate period of retention. This storage must ensure records are trustworthy,

accessible, readable overtime, and protected. This category is only for records as identified by the Records Retention Schedule.

The “value management" approach addressed both the business value and records value of information. This approach allows unnecessary information to be removed from the active email system (and later. from file shares and collaboration environments.) In doing so it would improve server performance and reduce storage costs while providing a system and procedures that satisfy legal requirements.

**Data Analysis**

A pilot study was conducted during the planning of the process. The results were evaluated and used for the analysis. Brainstorming session was conducted that included all the subject matter experts, stakeholders, vendors, business analyst and other project team members. All the requirements were gathered and were properly analyzed. In order to assist in assessing the company engaged a consulting group in gathering requirements

**Budget**

Table 2 represents the estimated costs for proposed solution (+-50%).

Table 2: Estimated Costs

| SNO | Capital | Year 1 | Year 2 | Year 3 |
|---|---|---|---|---|
| 1 | Licenses (email management software) | $1,500,000 | $1,500,000 | $750,000 |
| 2 | Hardware (servers and server-related software) | 250,000 | 500,000 | 250,000 |
| 3 | Implementation (external resources) | 250,000 | 250,000 | 150,000 |

Table 3 represents the expenses for implementing the project.

Table 3: Expenses

| SNO | Capital | Year 1 | Year 2 | Year 3 |
|---|---|---|---|---|
| 1 | Internal Resources | 200,000 | 200,000 | 200,000 |
| 2 | Design and Implementation Consulting | 250,000 | 200,000 | |
| | Totals | $2,650,000 | $3,350,000 | $2,100,000 |
| 3 | Change Management (including training) | 200,000 | 400,000 | 150,000 |
| 4 | Maintenance | | 300,000 | 600,000 |

With the approval to proceed and place as an item into the year budget, the project team undertook the following steps:

- Continue to track resource time to the existing corporate systems project code until a new budget item/project code is established or cut.
- Begin technical sessions with all three vendors to understand the architecture to flesh out hardware sizing and any additional costs in regards to software.
- Have procurement start their research to gain a clearer understanding of the true licensing and professional services (design team) costs.
- Estimate staffing (FTEs) for ongoing operation of this technology/ product/ service and identify which part of the organization would own it.
- Follow-up vendor reference accounts.
- Gather all information captured.
- Regroup senior management on our findings, a recommended vendor and a more complete and accurate costs/risks/schedule to be entered into the year budget.

**Timeline**

Initially the proposed timeline was as below:

*PART 1 (New Solution):* Assuming that this would be a priority project and resources are appropriately dedicated, it will take 8 to 10 months to build development, QA and production environments. Another 3 to 6 months will be required to move the enterprise onto the new solution. Professional services are required for detailed system, perhaps with temporary resources to assist with the initial infrastructure workload resource curve.

*PART 2 (Legacy Sunset/Retirement)* Once the company is on the new solution, it will be important to immediately retire, migrate, or otherwise safeguard the legacy repository (essentially comprised of EmailXtender software and Centera storage hardware) to avoid risks that the system fails and the information is no longer available. EMC has stated the product will require an extended support agreement in 1 QTR for the EmailXtender software. The current support contract is $120,000 per year. Below is the gnat charting the represents the timeline for project.

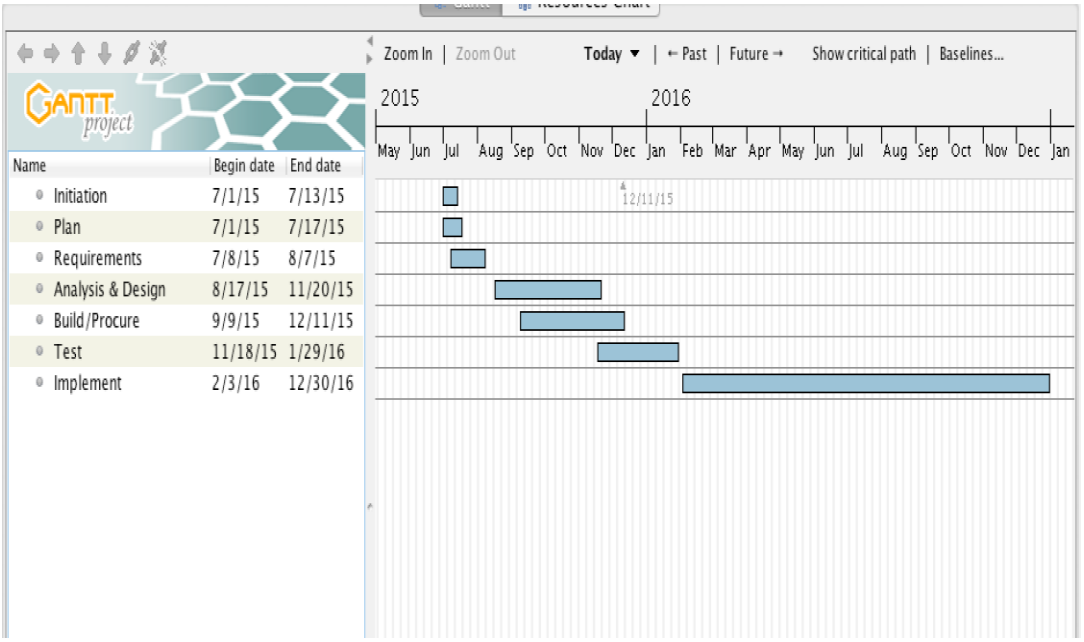


Figure 5: Timeline

**Summary**

The chapter in detail explains about the approach to the project. It is a road map for how the project was conducted. It gives a brief introduction about data collection process. Further it explains in details about the budget that was involved during the product and the timeline for the project implementation and deployment.

# Chapter IV: Data Presentation and Analysis

## Introduction

This chapter will explain in detail what is done during data analysis and it explains in detail the exact procedure and the steps. And data collected will be presented.

## Data Presentation

The consulting company assisted in refining and finalizing the company's requirements for email Management, records management and e-discovery, which was based upon the foundation of the Generally Accepted Record-keeping Principles (GARP) recently released by Association of Records Managers and Administration (ARMA International). These principles establish a solid and defensible architecture for a records management program including related technologies. Below is a list of requirements categories and associated categories including the RFP.

a. *System Requirements:* The system requirements explains the requirements of the solution to coexist in the existing companies environment, and to enable the value handling of email based on its business, legal and regulatory values. It included requirements for:

- Integrated of proposed solution component
- Integration with existing companies software
- Management of the requirements for systems control and monitoring
- Support for the messaging environment and all devices
- Ability to handle disaster recovery and disposition requirements

- Granular explanation of Zone 3 information management
- Constraints of each zone
- Characteristics of each zone

b. *Administration:* The administration section explains the requirements of the solution for software application administration including general configuration requirements, configuration support of company's policies and administrative monitoring and auditing. This section includes requirements for:

- *System Configuration*:
  - Configuration of taxonomy and associated record category rules and metadata
  - Distribution of application administration
  - Configuration of information access, restrictions and ownership
  - Centralization and distribution of all configurations
- *Retention and Disposition*
  - Creation of fixed and event based retentions
  - Disposition of primary objects and related data
  - Certification of destruction
- *Security*
  - Development of security guidelines and classifications
  - Assignment of access rights

- *Privacy*
    - Lifecycle management of Personally Identifiable Information
    - Operation of multiple jurisdictions
- *Monitor and Audit*
    - Application of administration monitoring and corrective actions
    - Reporting

c. *End User Acceptability:* This section explains the requirement of the solution from the end user perspective, including their typical interaction with the solution. These requirements include:

- *Setup*
    - Ease of set up
    - Personalization and customization
- *Declaration and Classification*
    - Declaration and classification workflow
    - Automation and end user assisted classification
- *Search and Retrieve*
    - Single search interface
    - Retrieval of original formats and creation of copies
- *Usability*
    - Commonality for end user in regards to look and feel across applications

- Movement to repository workflow
- Management of user dashboard

**Data Analysis**

*Score card analysis.* The company and consulting firm reviewed details, RFP responses from seven vendors. After the evaluation, three vendor finalists were each invited to participate in a 4-hour vendor demonstration of their capabilities to meet company requirements. A scorecard was utilized to document capabilities, including strengths and weaknesses, during the vendor demonstrations. A summary of scoring cards is included in Table 4,

Table 4: Score Card Analysis

| SNO | Evaluation Category | IBM | Open text | Autonomy |
|---|---|---|---|---|
| 1 | System requirements | 240 | 221 | 225 |
| 2 | Administration | 199 | 184 | 156 |
| 3 | End User Functionality | 127 | 88 | 95 |
| 4 | Litigation Readiness | 196 | 167 | 202 |
| | Grand Total Scores | 762 | 660 | 678 |

*Recommended solution.* Preliminarily, IBM has demonstrated the most complete functional solution with Open Text and Autonomy as runner-ups. An analysis was done on how the IBM solution addresses the eleven business problems with the existing email archive and archiving process that were identified as the genesis of the project. This analysis resulted in a favorable rating for the proposed solution in addressing the original eleven business problems.

After the scorecard analysis was done there were certain assumptions and constraints where further analysis was done to insure IBM can provide the most functional solution.

*Assumptions/constraints.*

1. Vendors costs in the RFP are high level, likely accurate only to a +-50% level. Once all of the other issues here are addressed, Procurement will undertake a final round of pricing negotiations to achieve more favorable pricing. At the same time, further research by the core team and consulting company will more accurately estimate the total internal and external resources.
2. During this sizing, IT also requires a deeper technical review of all three of the vendors to evaluate the compatibility of the various parts of their solutions with existing company's infrastructure, as well as company's ability to support the new.
3. During the technical review, an investigation of what Microsoft Office/Outlook 2007 or 2010 dependencies the three finalists may have is required. These dependencies could affect initial solution costs and complexity and the size and scale of subsequent user change management efforts.
4. Other specific concerns about how IBM is meeting companies requirements also need further research, including, the uses of content collector: what other additional products are required for e-discovery

purposes; and what the dependency will be on the small third party IBM partner, Integro, to meet companies needs now and in the future.

5. Any change to how end users will be interacting with email will require diligent final design of the user experience; rigorous acceptance testing; and a significant change management effort to ensure successful implementation and assimilation into the normal course of business.

**Summary**

This chapter explained in detail how the data analysis was done and the steps included. It further explained in detail about the data presented from different vendors and what were finally chosen.

# Chapter V: Results, Conclusion, and Recommendations

## Introduction

This chapter is the conclusion of the document. It gives a detailed understanding of what was achieved by the implementation of the project and conclusion of the project. It also suggests certain recommendation on how the project could be even more successful for the further analysis.

## Results

Email archive management and retention was achieved to a successful level as per the desire of the business. It answered all the possible questions that were aroused at the beginning of the project. The retention is now successful until 2 years. End user acceptability was obtained to high satisfactory level. The project questions and detailed explanation of each question were addressed as follows:

1. *How will the current solution make e-mail records compliant to other supporting software?*

After the implementation of the project email records are compliant than before. The new solution supports application based email such as Lotus notes email. The emails generated from application were able to be classified and go into the archive. The new system is capable for messages and their attachments, calendar entries, contacts, notes, to dos, and journals. The management capabilities included retention, disposition, record disposition and classification. It has the capability of filing emails and attachments not only when directly or remotely connected, but also when working offline. When reconnected to the network,

synchronization will be automatic. Each zone must store the email according to the category and lock them down for the retention period.

2. *How does the project ensure personal identifiable information is secure?*

The new solution prevents the use of personal email archives. It provides the ability to import personal email archives into a central repository. Data in desktop, email application or record repositories and in any repositories have a synchronized deletion function such that data is purged from repository during the same business day as it is purged from the active email backup. The new solution has the ability to centrally configure individual user profiles that are preloaded with the record categories relevant to the function/role of the user. These are either pushed to the user or selected as favorite by the user or both.

3. *How long would the emails be retained?*

The emails received or created are default to Zone 1. Such emails must be moved out of the inbox and retained for about 15 months and then deleted after review, unless moved by the user to ‘Zone 2’ or ‘Zone 3’ prior to the end of that time. Email moved to ‘Zone 2’ must be subject to space and time limits with assisted deletion after the prescribed period unless declared as record and moved by the user to ‘Zone 3’ prior to the end of that time. Must be able to dispose of messages in the normal course of business. ‘Zone 2’ will have a maximum 24-month retention behind the scenes.

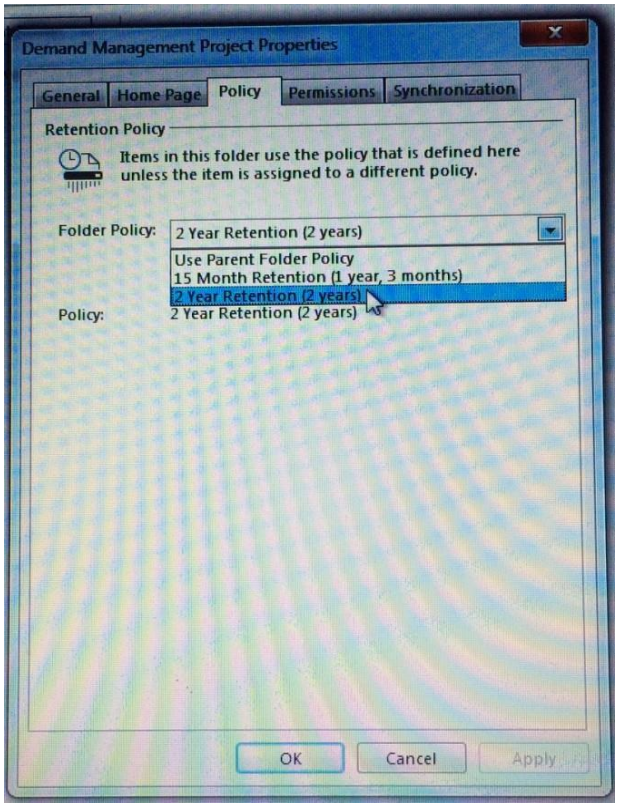


Figure 6: Retention Options after the Implementation of Project

4. *How capable is the retention schedule to be acceptable by end user?*

The retention schedule was completely acceptable by the end user. The new solution is capable of deleting obsolete and redundant emails. This solution eliminates the need for personal email archives. It provides a practical location for end users to store and protect the integrity of records prescribed by company retention policies.

5. *Can we have another simplified solution?*

After the requirements were gathered and a solution was proposed it seemed to be well in place. But by the end of the project there were other possibilities of

having more simplified and precise solution. Value management approach worked very well but this could be more precise.

Apart from these questions, the project further answered the solution for the entire eleven-business problem that existed before the implementation of project. The solutions that came up by the end of the project were:

1. Replace archive with content management repository.
2. Save the cost of an upgrade or extended support of an ailing and outdated solution.
3. Delete, in the normal course of business, an estimated 50-70%of emails flowing through the email system that are transitory in 30 days after creation or receipt.
4. Improve accuracy of tools for searching the lower volume of stored emails, reducing the data sets returned and as a result bringing down the cost of e-discovery.
5. All email must currently be held indefinitely since preservation obligations are not known.
6. Disposing of emails per retention policies to ensure emails for legal and regulatory purposes are retained for mandated internal/external time requirements.
7. Having emails stored in both EMC Legato and in PST files, and also having larger amounts of data to search than needed increase costs.

8. Store long-term email in managed repositories by content per company retention policy.
9. Provide future capability to manage IM and voice records like email.
10. Retain and dispose of emails per retention policies to ensure emails needed for legal and regulatory purposes are retained for mandated internal/external time requirements.
11. Discontinue use of PST files to keep long-term emails.

**Conclusion**

The email archive management using IBM tool has proved to be a great success with high efficiency. The achievements from the project were as expected. The assign policy button works as expected. The save as record works as expected reach its goal for security, search and templates. The project provides thee users more efficient way of retaining the emails.

It provides the company to have a great ability to:

1. Implement legal holds on email.
2. Search, preserve and collect email.
3. Reduce exposure to litigation, investigation and audit.
4. Decrease catastrophic loss of information.
5. Enhance credibility for compliance with regulations and customers.
6. Improve defensible management of email.
7. Improve information access, confidentiality and privacy protection and security.

8. Reduce storage and resource cost for managing retained email.
9. It gave a solution for all the business problems and solved according to the severity.

**Recommendations**

The project was very well planned and properly implemented. Though it is properly done one recommendation was to leverage out of the box ECM capabilities for executing project. All the capabilities were considered that were very hard to analyze and this made certain steps to be reconsidered.

## References

Brogan, M. P., & Vreugdenburg, S. Y. (2008). You've got mail: Accountability and end user attittudes to email management. In *Proceedings of International Conference on e-Government* (pp. 63-70). RMIT University Melbourne Australia.

Cameron, S. A. (2011). *Enterprise content management: A business and technical guide*. Swindon, UK: Charlesworth Press.

Floyer, D. (2009, March 21). *Email archiving*. Wikibon. Retrieved from http://wikibon.org/wiki/v/Email_archiving.

Kidda, T., Grego, B. Caplinger, C., & Houberg, R. (2011). *SharPoint server 2010 enterprise content management*. Birmingham, UK: Wrox Press, Ltd.

Message Labs. (2009, September 25). *Spam spikes: A real risk to your business*. White Paper. Author.

# Appendix

## Analysis and Comparison between Different Vendors for the Purpose of Choosing the Right One

### 1. System requirements

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 1.1 | All systems comprising the solution and their components must be integrated. The system must integrate within the current companies architecture | | 2 | 1 | 1 |
| 1.2 | All applications and repositories must be managed under the control of companies IT | | 2 | 2 | 2 |
| 1.3 | Must support de-duplication of message, attachments, and files while allowing multiple classification to the same electronic document or email | | 2 | 2 | 2 |
| 1.4 | The use of personal email archives must be prevented | These should include Email should not exist outside of centrally controlled archive environment There must be single repository for email archiving | 2 | 2 | 2 |

### 2. Zoned Information Management

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 2.1 | The solution must provide functionality to allow users to manage files using the proposed '3 Zone' model | These include, After a user receives an email, the system will require the user to classify the email when they move it out of the box | 3 | 3 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 2.2 | Must have a consistent classification structure and process for email with the ability for users to create subfolders in all zones | | 3 | 2 | 1 |
| 2.3 | The document record classification of company records must be visible to internal companies. External recipients must not have any knowledge of the record status | | 2 | 2 | 0 |
| 2.4 | Must be able to support disposition of information according to the retention policy and record retention schedule in Zone 3 repositories | | 3 | 3 | 3 |

## 3. System Configuration

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 3.1 | Should provide for a consistent taxonomy and glossary of terms to be defined, maintained and enforced through a mechanism transparent to the user | | 2 | 2 | 4 |
| 3.2 | The company record categories and controlled terms for individuals, business units, work group functions and geographic regions should be available under configuration | | 2 | 2 | 4 |
| 3.3 | Ownership of information should be explicit and linked to retention schedule; where appropriate must also identify system of record | | 2 | 2 | 4 |
| 3.4 | Should identify group and individual information ownership | | 2 | 2 | 4 |

## 4. Retention and Disposition

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 4.1 | Must adhere to the records retention schedule and written policies to allow the end user to manage email and attachments | | 3 | 2 | 6 |
| 4.2 | Event information is captured when the retention is based on an event, or an event plus a period of time. The event name should be captured or pre-configured at the time of classification so that the future event trigger can be associated with the record | | 3 | 3 | 9 |
| 4.3 | The event dates can be populated by an end user or administrator preferably by an automated method | | 3 | 2 | 6 |
| 4.4 | Must allow the disposition of transient, obsolete and redundant data automatically after user notification | | 3 | 2 | 6 |

## 5. Security

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 5.1 | Must support the application of one or more security classifications at the document or email level | These should include:<br>End user access requirements:<br>Ability to mark emails as not being viewable to end user<br>Ability to restrict access only to one's email<br>Ability to alert on user mass deletions of emails from their "view" | 3 | 2 | 6 |

| | | | | | |
|---|---|---|---|---|---|
| 5.2 | Must have the ability and means to define and support a security and access model | | 3 | 2 | 6 |
| 5.3 | Must enable desktop and email applications to support dual key encryption functionality for email and attachments at rest | These include:<br>Microsoft rights management services interoperability | 1 | 2 | 2 |
| 5.4 | Company data must at all times be adequately protected from inappropriate access by employees or contractors | | 3 | 2 | 6 |

## 6. Privacy

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 6.1 | Should support the appropriate protection and management of personally identifiable information throughout its lifecycle | | 2 | 2 | 4 |
| 6.2 | Should support multi-jurisdictional personal privacy and business secrecy obligations regarding access, storage, use, processing, disclosure and transfer of electronic information | | 1 | 2 | 2 |
| 6.3 | Should enable the disposition of personally identifiable information after its useful life in accordance with the prescribed disposition rules, as only the minimum necessary personally identifiable information will be retained pursuant to company policy | | 2 | 2 | 4 |
| 6.4 | European email servers should be located in an approved location. This will be the responsibility of Company | | 3 | 2 | 6 |

## 7. Monitors and Audit

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 7.1 | Must provide a set of work processes allowing a designated administrator to monitor and correct, maintain and update event and hold dates, review records for scheduled disposition actions and execute disposition and destruction | | 3 | 2 | 1 |
| 7.2 | The system must provide reporting capabilities | | 1 | 2 | 1 |
| 7.3 | System will include Adhoc reporting capabilities | | 2 | 2 | 2 |
| 7.4 | Must identify email and document declared as company records | | 2 | 2 | 1 |

## 8. Setup

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 8.1 | Must be easy for end user to setup client MS Office, desktop application and email applications | | 3 | 3 | 1 |
| 8.2 | The user must be able to modify or customize individual user profiles. | | 3 | 2 | 1 |

## 9. Declarations and Classification

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 9.1 | Must support additional classification by end users | | 2 | 2 | 2 |
| 9.2 | The recipient should be able to take normal actions on a message or electronic document that was classified as a record by the originator including classifying differently. When various recipients classify their copy of an email or document different from one another, the integrity of each record including metadata must be maintained | | 2 | 2 | 2 |
| 9.3 | Must support selective automatic classification of electronic records | | 2 | 0 | 2 |
| 9.4 | Must allow auto-completion or assisted classifications with end-user override to be utilized to guide user categorization of documents and email | | 2 | 0 | 2 |

## 10. Search and retrieve

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 10.1 | Must have capability to maintain and search through a single interface, electronic records in a manner that ensures timely, efficient, and accurate retrieval of needed information | | 2 | 2 | 2 |
| 10.2 | Must have the ability to search the message, full body, metadata and attachments | | 2 | 2 | 2 |
| 10.3 | Must allow emails and documents to be stored and retrieved in the original format | | 2 | 1 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| 10.4 | Must have the ability to individually view or retrieve multiple emails or files from a single search | | 2 | 2 | 2 |

## 11. Usability

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 11.1 | Must have a common end user approach with minor deviation from the existing Exchange interface. Users should not be required to initiate another application | | 2 | 2 | 0 |
| 11.2 | Moving of data to the repository must be automatic and transparent upon user initiation | | 3 | 3 | 3 |
| 11.3 | User will have the ability to modify access by document, message or folder within the constraints of the security classification model | | 2 | 0 | 2 |
| 11.4 | A dashboard or statistics should be readily available to assist the end user in managing all zones of their active mail file | | 2 | 0 | 2 |

## 12. Identification

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 12.1 | Must be able to execute a variety of search techniques against data attributes under the control of the system to identify objects containing potentially relevant information | | 2 | 1 | 3 |
| 12.2 | Must permit loading filter terms, date ranges, user identification criteria, enabling a quality assurance check against this information, and then kicking off an automated job that will identify responsive objects | | 2 | 1 | 3 |
| 12.3 | Must be able to identify objects containing a wildcard term | | 2 | 2 | 2 |

| | | | | | |
|---|---|---|---|---|---|
| 12.4 | Must support functionality to search by custodian, matter, full text | | 2 | 1 | 3 |

## 13. Preservation and Legal Hold

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 13.1 | Must support the ability to preserve email objects in place so they are not accidentally or deliberately altered or deleted | | 2 | 0 | 0 |
| 13.2 | Must have the ability to place multiple legal/HR holds on emails by sender/recipients, date range, keyword, AD organizational unit and records classification that will not be subject to purging rules | | 2 | 2 | 2 |
| 13.3 | Should have the ability to track what was preserved and when it was placed on hold | | 2 | 2 | 2 |
| 13.4 | Must support functionality to lift a litigation hold | | 2 | 2 | 2 |

## 14. Collection

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 14.1 | Must be able to export all data identified as potentially responsive | | 3 | 3 | 3 |
| 14.2 | Must include functionality to export in their native format objects identified through the various search techniques identified above | | 2 | 1 | 2 |

## 15. Report and Audit

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 15.1 | Must include processes that are automated, capable of being run via interface or load file, and which generate logs or other reports that can be monitored for compliance with quality assurance criteria satisfactory to company legal team | | 2 | 2 | 2 |
| 15.2 | Must maintain an audit history per object, custodian and matter | | 2 | 2 | 2 |
| 15.3 | Must be able to maintain and report chain of custody information | | 0 | 0 | 0 |

## Workflow

| SNO | Description of Requirement | Further description | IBM | Open text | Autonomy |
|---|---|---|---|---|---|
| 16.1 | Must support the electronic discovery workflow activities, including discovery process workflow, ESI data map and inventory maintenance | | 9 | 6 | 6 |