



100 Year Archive Requirements Survey

January 2007



100 Year Archive Requirements Survey

January 2007

Authors:

- Michael Peterson** - Chief Strategy Advocate for the SNIA's Data Management Forum, President, Strategic Research Corporation
- Gary Zasman** - Co-Chair 100 Year Archive Task Force, WW Practice Director, Network Appliance
- Peter Mojica** - Co-Chair 100 Year Archive Task Force, VP Product Strategy and Management, AXS-One
- Jeff Porter** - Chair SNIA Data Management Forum
Senior Technologist, IMSG Office of the CTO, EMC

The 100 Year Archive Task Force is a program sponsored by the SNIA's Data Management Forum. The Task Force operates an online portal at www.snia-dmf.org/100year



Copyright © 2007 Storage Networking Industry Association (SNIA)

All rights reserved. All material in this publication is, unless otherwise noted, the property of the SNIA. Reproduction of the content, in whole or in part, by any means, without proper attribution given to the publisher, is a violation of copyright law. This publication may contain opinions of the SNIA, which are subject to change over time. The SNIA logo and the Data Management Forum (DMF) logo are trademarks of the SNIA. All other trademarks are the property of their respective organizations.

Table of Contents

EXECUTIVE SUMMARY

- The Digital Crisis 1
- The Requirements Survey 2
- Survey Highlights 3
- Requirements: 4
- Follow on Work by the Task Force 4

INTRODUCTION 7

KEY FINDINGS

- Highlights 9
- Requirements for Long-Term Retention 12
- Peer Recommendations 14
- Task force Recommendations 15

RESEARCH PROCESS AND RESPONDENT PROFILE

- Survey Methodology 17
- Respondent Demographics 18

SURVEY RESULTS

- What does Long-Term Mean 23
- External Drivers 25
- Information Profile 31
- Storage Infrastructure 36
- Operations Satisfaction Study 40
- Requirements for Long-Term Retention 51

APPENDIX

- About the Authors 61

List of Tables

- Table 1 - What Are The Top Pain Points? 11
- Table 2 - Top Pain Point Examples 14
- Table 3 - View Of Top External Factors As Drivers by Organization 26
- Table 4 - How Do You Physically Migrate Data On Nearline Tape? 37
- Table 5 - How Often Do You Migrate Data Retained On Disk? 38
- Table 6 - Top Pain Points: 51

List of Figures

1 - The 100 Year Archive Task Force Objectives	6
2 - Study Objectives	7
3 - Key Findings	9
4 - Key Findings	10
5 - Requirements for Long-Term Retention (I)	12
6 - Requirements for Long-Term Retention (II)	13
7 - Peer Recommendations	14
8 - Task Force's Recommendations	15
9 - Organization Type	18
10 - In What Industry Vertical do you Work?	18
11 - Respondent's Geographic Location	19
12 - Respondent's Job Title	19
13 - Which Organization Do You Represent?	20
14 - What is Your Responsibility Relative to Long-Term Retention?	20
15 - What Does Long-Term Mean?	23
16 - What Does Long-Term Mean by Job Responsibility?	24
17 - External Factors Driving Requirements	25
18 - What Does Legal Risk Mean?	26
19 - What Does Compliance Risk Mean?	27
20 - What Does Business Risk Mean?	27
21 - What Does Security Risk Mean?	28
22 - What Does "Other Risk" Mean?	28
23 - Top External Factors Driving Retention Requirements	30
24 - Who Defines Requirements for Business Information?	31
25 - Who Defines Requirements for Compliance Information?	31
26 - Types of Information Retained the Longest	32
27 - Longest Retention Requirement	33
28 - How Much Information is Retained?	33
29 - Which Application's Information is Most at Risk?	34
30 - Which Types of e-Mail are Important to Retain?	35
31 - Percent of Organizations Retaining e-Mail for > 10 Years	35
32 - Where are Long-Term Records Kept?	36
33 - What Retention Period is Stored on Tape?	36
34 - What Retention Period is Stored on Disk?	38
35 - Satisfaction With - Programs to Reduce Cost	40
36 - Satisfaction With - Programs to Solve Migration & Readability	41
37 - Satisfaction With - Programs to Reduce Risk	42
38 - Satisfaction With - Legal Discovery Programs	43
39 - Satisfaction With - Organizational Structure	44
40 - Satisfaction With - Retention Methods	45
41 - Dissatisfaction With - Retention Methods	46
42 - Profile of Information State	47
43 - What is Needed to Cope with Changing State?	47
44 - Rate Business Requirements for Long-Term Retention	51
45 - What is Needed from Archive Systems to Assure Readability?	52
46 - How Important are These Solutions to Retention?	53
47 - What are You Doing About the Top Pain Points?	53
48 - Typical Recovery Failures	54
49 - Analysis of Recovery Failures	55
50 - Peer Recommendations	58

EXECUTIVE SUMMARY

The Digital Crisis

Large organizations faced with retaining and preserving huge amounts of digital information for very long periods of time are at the front edge of a troubling crisis. Digital information¹ is actually easier to lose than if it were on paper or film. It is one thing to manage a domain of digital records that an archivist can personally guard and shepherd. It is quite another to meet the archival challenges of today's enterprise data center. These data centers can be characterized as environments with petabytes of distributed information, high data growth rates, many facilities and many departments with uncoordinated responsibilities and requirements, and lack of business-level budget, interest, and focus on its archives. All these operating challenges are now compounded by high risk. Yes, risk – risk of failure and fines from legal discovery, compliance requirements, or security threats. Add to this, the risk of losing information that may be of great value to the organization and the picture looks pretty daunting.

Top Four Ways of Losing Digital Information

- Can not read it
- Can not interpret it correctly
- Can not validate its authenticity
- Can not find it

The digital crisis is exacerbated by time. In 10 years, 50 years, 200 years, which applications will still be around? What computer and storage system will be able to read old information, providing that it is not corrupted by then? Even finding a single piece of content and all the linked-objects that contain associated content amid trillions of distributed information objects is at best, a costly adventure. The problems are huge and here is the dilemma. Many standards and best practices exist today documenting the practices of preserving digital information. Yet, none of them address the core problems caused by inadequacies and inefficiencies in the supporting storage infrastructure.

The ‘two grand technical challenges’ of long term digital information retention are logical and physical migration. Logical migration is the practice of updating the format of the information into a newer format that can be read and properly interpreted by

¹ Now also being called by the legal & compliance community “electronically stored information” (ESI),

The Digital Crisis

- Risk of losing digital information over time
 - Growing cost and complexity of physical and logical migration
 - Overwhelming volume of digital information to preserve long-term
 - Increased legal, business, and security risk
-

The Grand Challenges

Solve Logical and Physical Migration

The Petabyte Problem

Migration is broken. Migration practices do not scale to meet the digital preservation requirements of the data center.

future applications or readers without losing the authenticity of the original. Physical migration means to copy the information to newer storage media to preserve the ability to access it and to protect it from media corruption. Best practices today require logical and physical migration every 3-5 years. Based on these practice standards, the real underlying challenge is how to scale migration capabilities while controlling cost. An organization that has 1,000 TB (a petabyte, PB) in its digital archive repository will have 50% more next year. In three years, they will need to migrate that first petabyte. In five years they will need to migrate 2.25 PB. How do organizations expect to do that and keep up with the growth, the cost, and the complexity? The answer is they can not. They will not². It is the contention of the 100 Year Archive Task Force that migration as a discrete long-term preservation methodology is broken in the data center. Today's migration practices do not scale cost-effectively and won't be done until a crisis erupts. This means that today's reliance on migration is taking us down a 'dead-end path'. Hear this clearly. Under these practice guidelines, the world's digital information is at great risk! New technological approaches are required that meet the legal, business, cost, and scalability requirements of the 'digital age' for long-term digital information retention.

The Requirements Survey

In September 2006, the SNIA's 100 Year Archive Task Force decided it needed a clear statement of business requirements to frame and bound potential technology solutions to the long-term digital information retention challenges of the data center. The plan was to design and conduct an online survey inviting a broad range of information owning and administrating professionals worldwide to participate and provide guidance. Knowing that many of these practitioners did not have large-scale data center experience, the research plan's primary goal and chief assumption were these:

- **Goal:** Determine requirements for the definition of best practices and solutions to the long-term digital information retention problems of the data center.
- **Research Hypothesis:** Current archive practitioner's experiences with multi-terabyte-size archival systems are adequate to define the business and operating requirements for petabyte-size information repositories in the data center.

This report summarizes the findings of the requirements survey. It was conducted online and respondents were solicited worldwide

The Research Goal

Determine requirements for the definition of best practices and solutions for the long-term digital information retention problems of the data center.

² Only 30% of the 276 respondents to this survey claimed they do migration today.

through SNIA's many alliances. Respondents self-selected (meaning they decided if they could answer the questions properly) and there were no incentives other than personal motivation to help. It was a passionate subject as individuals from 276 organizations responded and completed the survey over the three month period of November 2006 to January 2007. Respondents came from three principal disciplines: Information Technology (IT), Records and Information Managers (RIM), and Archivists. Additional participants represented Legal, Security, and business groups.

Survey Highlights

First, the research hypothesis was validated. The respondents were very insightful and provided opinionated and experienced advice. The survey was also very successful because it derived a clear set of requirements that the Task Force can use to guide its work. Here are some of the important highlights.

- The survey establishes clear validation that long-term retention needs are real and that many organizations have long-term requirements.
 - 80% of respondents declared they have information they must keep over 50 years and 68% of respondents said they must keep it over 100 years. (See page 33)
- Long-term generally means greater than 10 to 15 years – the period beyond which multiple migrations take place and information is at risk. (See page 23)
- Database information (structured data) was considered to be most at risk of loss. (See page 34)
- Over 40% of respondents are keeping e-Mail records over 10 years. E-Mail is not just a short-term problem. (See page 35)
- Physical migration is a big problem. Only 30% declared they were doing it correctly at 3-5 year intervals. The rest of the sample group is placing their digital information at risk. (See page 38)
- 60% of respondents say they are 'highly dissatisfied' that they will be able to read their retained information in 50 years. (See page 46)
- Help is needed – current practices are too manual, too prone to error, too costly and lack adequate coordination across the organization. (See page 40-46)
- Collaboration and classification were recognized as very important practices to get the organization working together setting requirements for the management of their information. This recognition reinforces the messages of the SNIA's Data Management Forum (DMF) in its market educational efforts for Information Lifecycle Management (ILM)-based practices. (See page 53)

The Survey

- Online, Quantitative Survey
 - 276 Organizations
 - IT, RIM, Archivists, Legal, Security, and Business Respondents
 - World-wide participation
-

- Reinforcing this point, only 35% of respondents agreed with the statement that their IT and RIM departments coordinate requirements for retention and preservation of the information they protect. (See page 44)

Requirements:

The goal of this research was to define requirements from the practitioner’s viewpoint for long-term retention solutions. It succeeded. The respondent’s requirement feedback was summarized into four categories corresponding to the classes of needs that solutions or best practices must address. This list now defines the ‘market requirements’ that will guide the work of the Task Force.

The Four-Faces of Requirements

- Business Drivers
 - Barriers to Adoption
 - Operating Practices
 - Technology Challenges
-

- **Accommodate the requirements of the critical business drivers** behind long-term retention by mitigating legal, compliance, business, and security risk as well as preserving the history of the organization forever.
- **Overcome the barriers inhibiting adoption** of best practices that range from the cost-effectiveness of solutions to stimulating collaborative efforts within the organization. Many of these requirements are organizational issues and fit the profile of best practices.
 - The most alarming barriers were the warnings that executive management does not really care and that there is no prestige in archive practices within the IT organization.
- **Improve operating practices** by providing better management tools, best-practices, job visibility, and education.
- **Solve the technology challenges** by:
 - Solving logical and physical migration
 - Solving the ability to scale the volume of information
 - Incorporating metadata into the archival repository
 - Including databases, email, and legacy information
 - Providing a full spectrum of information and data services core to the digital information repository that provide for classification, control, discovery, availability, protection, security, integrity, audit, forensics, non-repudiation, preservation, and permanent deletion

Follow On Work By The Task Force

The 100 Year Archive Task Force has big goals. Foremost on the list is to solve the technical challenges. Using the requirements developed from this survey, work is progressing on several fronts:

- Produce a ‘Reference Model’ for long-term digital information retention

- The Task Force is developing a reference model similar in format to the OAIS³ document, covering the storage domain portion of long-term retention. Building off OAIS, it will define storage architectures and services that provide a robust, scalable, long-term digital information repository. The plan is through technology to completely change the concept of physical migration.
- Solving logical migration requires new format standards and a means to motivate application developers to implement those standards. The 100 Year Archive Task Force plans to leverage the OAIS architectural model's concept of an "archival information package" in creating a storage container being called Self-Describing, Self-Contained Data Format, (SD-SCDF). It is planned that implementation of SD-SCDF will be enabled by integration with the XAM⁴ application-to-storage interface standard currently in development by SNIA.
- Market Education – the Task Force operates a web-site and a speaker's bureau and is presenting at events world-wide. This report will be showcased and broadly distributed to assist in elevating the importance of organizations paying attention to their digital assets.

Based on the findings of this research, information professionals can be hard at work today creating a collaborative relationship between all departments in the organization with the goal of setting requirements for information. In Information Lifecycle Management (ILM) terms, these are the first steps in implementing a comprehensive ILM-based practice; collaborate then identify, classify, and set requirements for information. The problems and the requirements have been clearly articulated in this report. It is your responsibility to convince your organization of the importance of protecting your information assets and to put into place collaborative practices to identify and classify your information and then set requirements so that IT knows the business requirements.⁵

100 Year Archive Task Force Projects

- Reference Model including glossary and physical migration
 - Logical Format Standard
 - Integration with ILM and XAM
 - Market Education
-

³ OAIS: Open Archival Information System, ISO Standard 14721:2002

⁴ XAM: eXtensible Access Method – a new standard in development by SNIA that will potentially provide a platform for application adoption of SD-SCDF. <http://www.snia.org/xam/home>

⁵ For more insights on collaboration, see the ARMA-SNIA paper, "Collaboration-New-Std-of-Excellence" October 2006 at www.snia-dmf.org

Respondent Recommendation

Remember that IT doesn't own the information. RIM, Legal, Business units and IT all have a part to play in the decisions applied to business records and should be sitting down at the table together.

If you want to contribute to this project, participate in the 100 Year Archive Task Force which is part of the SNIA's Data Management Forum. You can help guide this work and elevate its effectiveness. Solving the long-term digital information retention and preservation challenge is very important and the Task Force needs experienced participants from many disciplines because of the complexity of these problems. You can learn more including how to get involved at www.snia-dmf.org/100year.

Figure 1

100 Year Archive Task Force

- Objectives
 - Produce a “best practices for long-term digital information retention” reference model similar to the Sedona project or OAIS
 - Solve the storage-side of physical and logical migration
 - Teach ILM-based practices for long-term digital information retention

THE SEDONA GUIDELINES:
Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age

A Project of The Sedona Conference™
 Working Group on Best Practices for Electronic Document Retention & Production

September 2004 Public Comments Draft

Reference Model for an Open Archival Information System (OAIS)
 (ISO 14721:2002)

The Sedona Conference™

INTRODUCTION

In September of 2006, the SNIA's 100 Year Archive Task Force began developing this survey. Its purpose is twofold. First, to define market requirements to assure that the work the Task Force has underway properly supports those needs. This study focuses on the challenges faced by information owners and administrators, and maps those challenges to the technology needs of the data center. And second, to use the results of the survey to promote the importance of this project and conduct market education. Market education is one of the core charters of the SNIA's Data Management Forum.

Upon launching the survey, many objections from RIM and Archivist participants were received, complaining that the survey was not in touch with their needs. Once it was explained that the problems the survey is studying deal with the huge and exponentially growing problem of long-term retention in the data center, the response became supportive. This study is unique in that respect. It takes the problems and needs that many information technology and information management professionals have in their current organizations and practices and maps them to the data center problem.

What is that problem? How do you cost-effectively manage, and automate the preservation of petabytes⁶ of information forever. How do you keep up with an overwhelming volume of information that is growing at 50% to 100% per year? How do you migrate multiple petabytes (PBs) per year logically and physically? The answer is you cannot do it economically. The data center archive process and storage system is technologically broken. It doesn't

⁶ A petabyte (PB) is one-thousand terabytes or a million gigabytes

Figure 2

Study Objectives

- Study the business and operational requirements for long-term digital information retention in the data center
 - **Goal:** Determine the requirements for long-term digital information retention in the data center. These requirements are needed to frame the definition of best practices and solutions to the retention and preservation problems unique to large, scalable data centers
 - **Research Hypothesis:** Practitioner's experiences with terabyte-size archival systems are adequate to define the business and operating requirements for petabyte-size information repositories in the data center

The Migration Challenge

An organization that has 1 petabyte, (PB) in its digital archive repository will have 50% more next year. In three years, they will need to migrate that first petabyte. In five years they will need to migrate 2.25 PB.

scale. This is the frame of reference of the study and it is essential that you read it from that context.

Another objective of the study was to look at the problems of the data center from multiple perspectives. Participants from RIM, IT, Archivist, Security, Legal, and business groups were solicited through the SNIA's many members, regional affiliates, and association partners. The outreach efforts promoting participation in the survey emphasized the need for feedback from people responsible and experienced in retaining digital records for long-term periods of time.

Capturing feedback from these different viewpoints was important and added to defining a rich and complementary set of requirements. Perhaps the most important message that came from the variety of respondents was the urgent cry for collaboration between the different disciplines. No single functional group can solve a cross-organizational problem. This finding is supportive of the SNIA-ARMA Task Force⁷ work, and the programs the DMF and ARMA are progressing to encourage collaboration between information owners and administrators.

The task of developing and conducting the online survey, conducting the marketing program to obtain respondents, analyzing the survey responses, and writing the survey report was led by Michael Peterson, President of Strategic Research Corp. and currently the Chief Strategy Advocate for the SNIA's Data Management Forum. Members and leaders of the 100 Year Archive Task Force and the SNIA Data Management Forum contributed to the effort. The survey was run as an open, online survey with self-selecting respondents from the IT, RIM, Business, and Archivist communities over the three month period of November 2006 to January 2007. A total of 276 individuals from a wide variety of organizations responded and took the 63 question survey.

The Survey

- Online, Quantitative Survey
 - 276 Organizations
 - IT, RIM, Archivists, Legal, Security, and Business Respondents
 - World-wide participation
-

⁷ See the paper , “*Collaboration-New-Std-of-Excellence*” October 2006 at www.snia-dmf.org produced by the SNIA-ARMA Task Force

KEY FINDINGS

Highlights

All of the questions asked in this survey produced important, usable findings applicable to better understanding the operating practices and requirements in this field. But, some are more pertinent than others and need emphasizing. This section of the report highlights those results.

Figure 3

As indicated by the list of findings presented in Figures 3 and 4, respondents are struggling with many problems in long-term retention. The needs are great and the problems are real. There are many expert practitioners in the respondent pool who will hopefully help guide this Task Force as it progresses its work.

Key Findings

- The problems of logical and physical retention
 - Practitioners are struggling – information is at risk long-term
 - Problems are real and generally understood
- Long-term generally means over 10-15 years.
 - IT can manage to migrate and retain readability for about this long. For longer periods, processes begin failing, become too costly, and the volume of information becomes overwhelming.
- Long-term retention requirements are real.
 - Over 80% of organizations reporting have a need to retain information over 50 years and 68% report a need of over 100 years.

“This is the problem with ‘Digital Archive’, you are not thinking long enough into the future.” (Source: Respondent)

Throughout the survey, important background questions were asked such as “what information-types are you retaining”, “how much data are you storing”, and “what does long-term mean to you?” The answers are profound. Long-term, by consensus, came out to be anything beyond 10-15 years because that is the time-frame beyond which they begin to lose control of logical and physical migration.

In addition, respondents identified that legal, compliance, business, and security risk, along with the fear of losing critical and historical records are driving the long-term retention of ever increasing amounts of digital information. They also verified that they have many real problems and are not confident in the “art” of preserving digital information for the long-term.

Long-term digital information retention is hard, complex, and affects the entire organization. Some of the more important and revealing results were these:

- Many more organizations than expected have a long-term problem. 80% of the 276 organizations reported a need in excess of 50 years.
- Database information was considered most at risk. (This problem is not limited to unstructured information or email.)

- Over 40% of respondents are keeping email records over 10 years.
- Physical migration is a big problem. Only 30% declared they were doing it correctly at 3-5 year intervals. The rest of the group is placing their digital information at risk.
- 60% of respondents say they are highly dissatisfied that they will be able to read their retained information in 50 years.
- Help is needed – current practices are too manual, too prone to error, too costly and lack adequate coordination across the organization.
- Collaboration and classification were recognized as very important practices to get the organization working together setting requirements for the management of their information. This recognition reinforces the messages of the SNIA's Data Management Forum (DMF) in its market educational efforts for Information Lifecycle Management (ILM)-based practices.

Figure 4

“The driving force for a true Archive is the preservation of the history of the organization for hundreds of years. Your survey does not address the historic needs. It is a matter of preserving for future generations the important facts (records) of an organization. This is the problem with 'Digital Archive', you are not thinking long enough into the future.” (Source: Survey Respondent)

Key Findings

- Respondents declare “Help is desperately needed.”
 - Current practices are too manual, too dependent on human intervention, too costly, too prone to error, and not adequately coordinated across internal information owning organizations.
 - 58% of respondents strongly agree that “IT professionals responsible for long-term retention do not know the ‘requirements’ for the information they protect”

*“Remember that IT doesn't own the information.
RIM, Legal, Business units and IT all have a part to play in the decisions applied to business records and should be sitting down at the table together.”*
(Source: Respondent)

The study concluded by asking some important summary questions about the respondent's organizational attitudes. The results illustrate two important points and potential barriers. First, the need for help is recognized. Most respondents talk about the constant investigation they are doing to try to stay abreast of technology developments. In addition, there is recognition of the complexity and cost of long-term retention practices. Second, the desire to collaborate is strong, as well as the recognition that senior management must be committed to solving the problem. The common paradox here is that archive information is important to an organization, but recognition for managing the archives is often low on the professional ‘totem-pole’. (This

complaint tied for the second highest “pain-point” in the survey.) Support, commitment, and professional recognition are issues that must be addressed.

Recommendations? Get RIM and IT at the same table. Create a relationship. Both need to be included up front to develop solutions that will work on both sides. Involve senior management early in the process and be sure there is a common goal outside the influence of technology enthusiasts. Collaborate and rely on standards and good practice. Communication and visibility about the issues of compliance with the agency's RIM policies and practices and enforcement of those policies (are needed). Upper management advocacy (is essential). (Source: Survey Respondent)

Respondent Recommendation

Collaborate: rely on standards and good practices and get management commitment

These points are amplified by the following list of “Top Pain Points”, and the following list of requirements. These illustrate the broad nature of the problems encountered by the functional groups represented in this study, as well as their view of the relative importance of each pain point. Long-term digital information retention is a complex problem that requires humans interface with numerous technology practices.

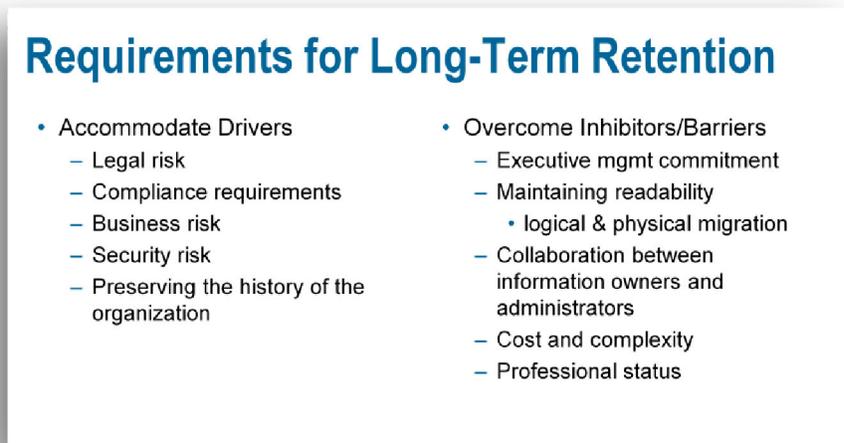
Table 1
What are your top pain points in long-term digital information retention?
 (Percent of Responses, Multiple Responses Allowed, n=276)

Media Migration	12%	Classification	8%
Maintain Readability	10%	Lack of Collaboration	7%
Technology Obsolescence/Upgrades	10%	Discovery & Deletion Difficult	6%
Lack of Business Support/Commitment	10%	Too Many/Legacy Formats	6%
Cost	9%	Lack of Expertise/Discipline	5%

Requirements for Long-Term Retention

The principal goal of this research is to define requirements from the practitioner's viewpoint for long-term retention solutions. A lot of valuable feedback was received. In tabulating and organizing it, four categories stood out: business drivers, inhibitors, operating practices, and technology problems. The principle is that all solutions and best practices that the Task Force defines need to satisfy the market requirements. Looking at them by category helps understand which approaches to take.

Figure 5



ACCOMMODATE DRIVERS: Meeting the needs of business drivers requires control over many important services throughout the lifecycle of millions of information objects. Control implies management of services and practices, including⁸ services such as these: classification, eDiscovery, protection, privacy, availability, integrity, auditing, preservation, and permanent deletion. The concept is to instrument, automate, measure, and mitigate risk of violating the requirements.

INHIBITORS AND BARRIERS: This category addresses needs practitioners have to overcome inhibitors or barriers to adoption. As highlighted in Table 1, lack of commitment and collaboration are very highly rated pain points. Lack of senior management commitment is a common organizational barrier which inhibits support and budget and leads to failure. Similarly, collaboration between the various functional groups is critically important and has to be encouraged through ongoing educational activities. Other items on the list include parameters that overlap other requirement categories because in the respondent's mind, unless they get solved, they are barriers. The success of practices and technology solutions for long-term retention will be gated by how well those solutions help the practitioners in the field overcome these barriers and are easy to adopt.

⁸ This list of services is not all-inclusive. eDiscovery means 'electronic discovery', a service analogous to content search of digital objects.

Lack of management commitment is a major concern. Consider this respondent's perspective:

As we complete the move to Electronic Medical Records, clinical files are being automatically generated. This initiative is driven by federal and health care mandates and is progressing pretty well for our field. The business side of the operations is very far behind the curve. We're a 100-year-old institution with NO records management structure. It's very scary to me that the administration is so cavalier about business records. (Source: Survey Respondent)

Respondent Concern

Management
Commitment

OPERATING PRACTICES: The third category addresses requirements that affect practitioner's ability to perform their jobs. The responses range from organizational to technology. A number of the top pains keep showing up such as migration, collaboration, and commitment. Unique to this list are criteria like reducing operating costs and better management tools. Across a number of questions, study respondents expressed their needs, satisfactions and dissatisfactions with long-term digital information retention. In doing so they articulated the observation that there are no 'silver bullets' available. Long-term retention is hard work as further illustrated by the examples listed in Table 2.

The distribution of (what is being stored) on disk must match the ongoing business value of the data – automatically. If (this does) not (happen), management is an unsolvable problem, since humans cannot keep up with the data onslaught. (Source: Respondent)

TECHNOLOGY ISSUES: The technology issues that respondents listed range from migration, to better tools, to better practices. Respondents included important requirements such as "including all information types", "including new and old legacy information", and "adding metadata to existing information". Here is a good checklist to measure potential solutions.

Requirements for Long-Term Retention

- Practitioner's Needs
 - Solve logical and physical migration
 - Solve technology obsolescence
 - Improve business commitment
 - Reduce operating cost
 - Better management tools
 - Better collaboration
- Technology Issues
 - Solve logical and physical migration challenges
 - Solve the scaling problem to keep up with the growing volume
 - Information classification
 - Include provenance & metadata
 - Dealing with growth & technology change/obsolescence
 - Include databases and email
 - Include legacy Information
 - Better discovery and deletion

Figure 6

Table 2
Top Pain Point Examples

-
- IT reluctance to destroy data
 - Risk of degradation
 - Migration issues
 - Conversion and migration costs and effort
 - No good long-term solutions
 - Unwillingness of top management to recognize problem
 - Business units & IT are unaware of the real retention requirements
 - Metadata capture
 - Risk of obsolescence
 - Poor/inconsistent metadata
 - Fluid technology
 - Organizational structure and leadership
 - Lack of coordination between IT & RIM
 - Technological obsolescence
 - Lack of Standards
 - Proprietary formats

Peer Recommendations

Figure 7

The last question of the survey asked respondents to summarize their experiences and make their own best practice

recommendations to their peers. The feedback was insightful as it added a human element to the discussion. While all the feedback is valid and important, with few exceptions these recommendations only address the organizational and human-factors side of the problem. It is now up to the storage and application providers to address the technology problems, while remain-

Peer Recommendations

- Recognize this is difficult
- Appreciate how important it is for the future
- IT doesn't own the information
- RIM, Legal, and Business units have a role in decisions and requirements
- Form a team to classify information and set requirements
- Metadata is important, Implement it in the front-end
- Stay with industry standards
- Get on top of your data retention needs now
- Get trained and stay on top of industry developments for long-term retention
- Get commitment from management and establish common goals

ing sensitive to the expert practitioner's requirements. This quote sums it up quite well.

“When using a digital archive, understand you will have a long hard expensive road to keep the records. You have to think about the ability of your great, great, great, great ... grandchildren being able to read and logically interpret what your history was...” (Source: Respondent)

Task Force Recommendations

By combining the respondent's feedback, the 100 Year Archive Task Force now has a set of requirements that bound the approaches to finding and adapting technology solutions. These requirements incorporate elements from all four categories of needs.

Storage Best Practices Requirements

- Assure digital information is being preserved for its retention period
 - Know and set clear business, legal, and compliance requirements and only retain long-term what is needed
 - Assure information is available, accessible, findable, private, secure, protected, and readable
 - Manage duplicates and versions
 - Provide for permanent deletion at the end of the retention period
 - Provide auditing capabilities and a periodic review process of what is being retained and its requirements
- Assure originals are authenticated and preserved
 - Store in an 'archive' format
 - Include metadata, provenance, and finding aids
- Put in place controls to handle
 - Hardware and software obsolescence
 - The ability to read and interpret information over its lifecycle
- Plan for logical and physical migration and automate management of the repository

Another dimension of the core findings of this study is that the storage industry cannot develop solutions to long-term retention in a vacuum. It is the Task Forces' opinion that effective solutions and practices must be multi-disciplinary and integrative with existing best practices and standards including:

Figure 8

Task Force's Recommendations

- Response to business drivers
 - Solutions & best practices that are developed have to be compatible with requirements for compliance, discovery, integrity, privacy, protection, etc.
- Inhibitors to overcome
 - The three technical inhibitors (migration, cost, & complexity) are essential elements of proper solutions
- Target solving the top storage-related problems
 - Physical & logical Migration
 - Integrating meta-data
 - Reducing management cost and reduce operating costs through automation
 - Keep information available, discoverable, protected, private & secure
 - Integrate with XAM, ILM, and other existing standards & practices

Requirements
Guidance

Multidisciplinary and integrative with existing standards and practices

- The RIM, Archivist, and IT practitioners, who will guide the approach and evaluate the solutions in each organization
- The application providers, who need to generate the correct metadata upon creation and package long-term formatted information objects into containers.
- Utilizing existing standards such as OAIS and best practices established by the Sedona Conference
- Integrate with existing and emerging storage standards such as SMI-S and XAM to enable automation and ILM-based practices
- Provide for consistency with established operating methodologies such as IT Service Management and Information Lifecycle Management.

In reviewing the survey from a solutions perspective, it is quite fascinating that no one talked about using solutions based on existing standards such as OAIS. No one offered a comprehensive solution to the many challenges they face. Instead, all based their current operating practices on the foundation of a continual search for better practices and complained about the magnitude of the problems they face. In essence, simply stating “We have all these problems and don’t have good solutions.”

This study confirms that the digital information storage industry has a crisis looming ahead. If it is not solved soon through innovative standards-based technology solutions, it will only get worse. As a classic example, the film industry’s Science and Technology Council recently endorsed an archival solution based on recording digital masters back to film as the only reliable preservation method available today.⁹ The question at hand is not if this trend will migrate to the data center. Many digital information objects now contain relationships, content, and links that can not be captured or portrayed by analog media anymore. It is no longer as simple as one respondent quipped, “*If you want to save it long-term, copy it to paper!*” We have only one viable option, solve the problems or prepare to lose the data.

⁹ “...Earlier this year, three companies received Science & Technology Awards for their work on archiving. Feiner and his Pacific Title team were among the winners. Their solution takes the data from a digital intermediate and turns it into three-color separation negatives. In other words, they take the digital movie and turn it into good old-fashioned film.” (Source: Variety International, “Digital proves problematic” David S. Cohen, April 20, 2007.)

RESEARCH PROCESS AND RESPONDENT PROFILE

Survey Methodology

The survey was online, collecting data from November 2006 to January 2007 drawing from an open, self-selecting audience. The survey was promoted world-wide through SNIA, ARMA International, and Society of American Archivists (SAA) association web-sites and newsletters. It was picked up by some additional friendly web-sites and discussed vociferously in several RIM and archivist bulletin boards. Interest was high and the debate stimulating.



Here was the introductory text used to help potential respondents, self-select.

“This survey is designed to capture business requirements for long-term digital information retention within the data center. Questions are asked about details of your operating practices and needs. This survey is anonymous to protect your ability to share information. However, we do need some basic profile information on you to be able to understand your responses in context with your responsibilities and to be able to 'cross-tabulate' the data. Please fill this section out completely. Please also note that it is better if you skip questions that you don't have accurate information about or don't apply to your situation - rather than guessing and potentially biasing results.

The 100 Year Archive Task Force operates from within the SNIA's Data Management Forum as a multi-agency committee working to define the storage requirements and best practices for long-term digital information retention. Storage systems are integral to long-term retention and new standards and best practices are needed to keep up with changing requirements and technologies.”

The survey consisted of 63 questions in many formats including open-ended “fill in the blank” questions so that unbiased results and respondent comments could be captured. The survey took about 30 minutes to complete. Not all respondents were able to answer all questions. This was an expected result considering the variety of backgrounds and range of experience. The survey questions ranged from operating practices, to in-depth details about application and capacity profiles, to questions about each functional group’s responsibilities and needs.

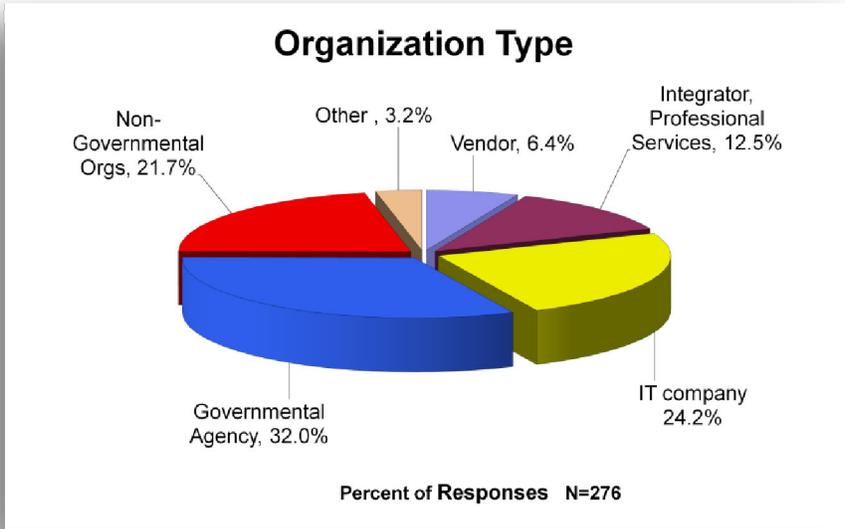
100 Year Archive Task Force Research Team:

- Michael Peterson, Chief Strategy Advocate for the Data Management Forum
 - Gary Zasman, Co-Chair 100 Year Archive Task Force
 - Peter Mojica, Co-Chair 100 Year Archive Task Force
-

Respondent Demographics

This section presents profile and demographic information on the 276 survey respondents and their organizations. Use it to understand the profile of the respondents and to assess the success of the survey in drawing in a large, diverse set of respondents.

Figure 9

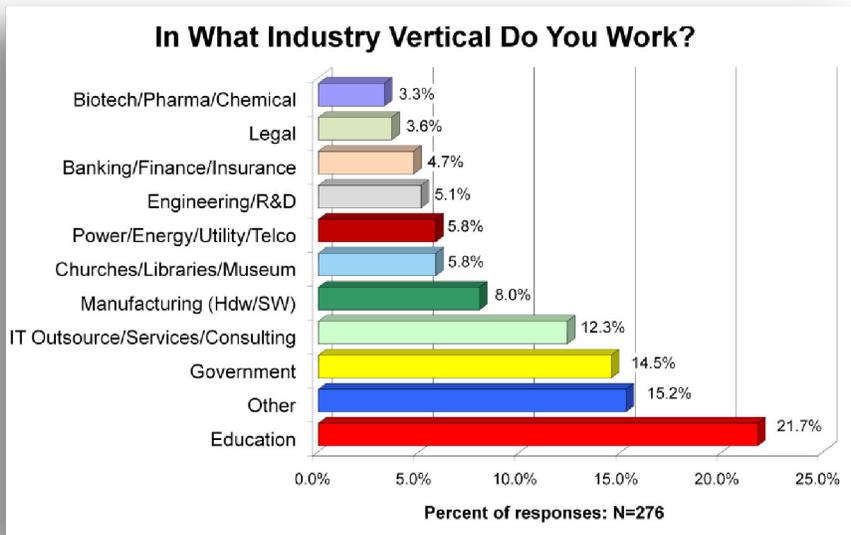


ORGANIZATION TYPE: The mix of respondents was in line with where long-term retention pain exists, governmental agencies, non-governmental organizations such as universities, libraries, and museums, and IT companies. The individuals representing the few vendors in the respondent mix were in professional services or IT outsourcing dealing with long-term retention and preservation problems for their clients.

INDUSTRY VERTICALS:

With RIM, IT, and Archivists as the dominant respondent it is no surprise that the leading verticals represented by their organizations are education, government, IT services, and places where archivists work, including Libraries, Museums, and Churches. 65% of the respondents represented a very broad spectrum of organizations which further validates the importance and relevance of solving the long-term retention issues.

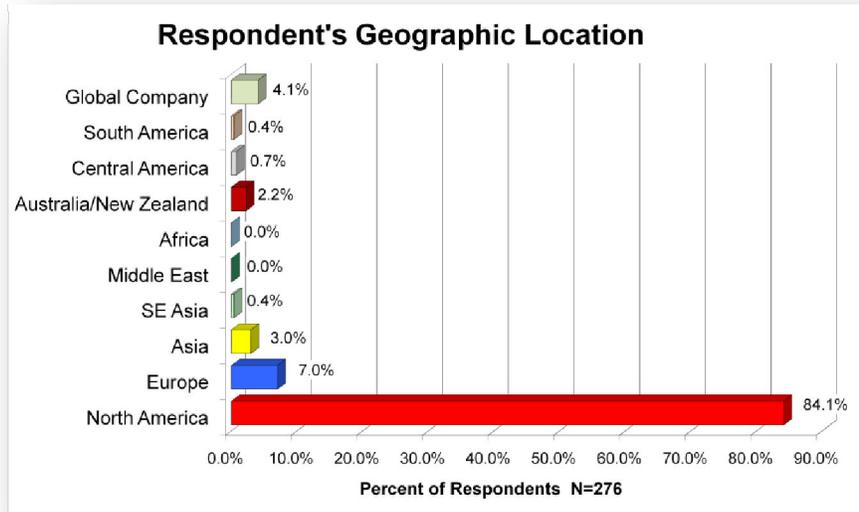
Figure 10



INDUSTRY VERTICALS: With RIM, IT, and Archivists as the dominant respondent it is no surprise that the leading verticals represented by their organizations are education, government, IT services, and places where archivists work, including Libraries, Museums, and Churches. 65% of the respondents represented a very broad spectrum of organizations which further validates the importance and relevance of solving the long-term retention issues.

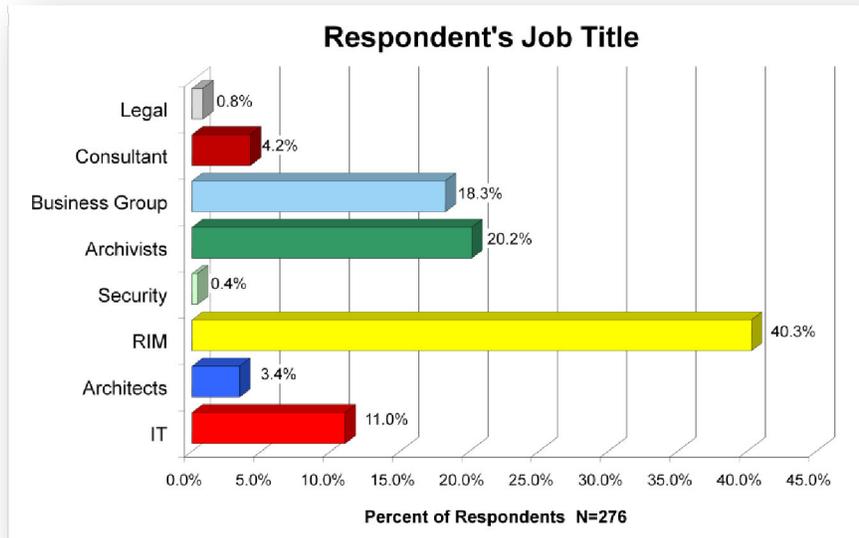
GEOGRAPHIC LOCATION: Most respondents were based in North America and Europe (91%) with a small percent from Asia and Australia/New Zealand and just a few from other geographies. This response provides more evidence of universal concern with the long-term retention problem as the survey was marketed actively in North America and Europe and very lightly in Asia and Australia.

Figure 11



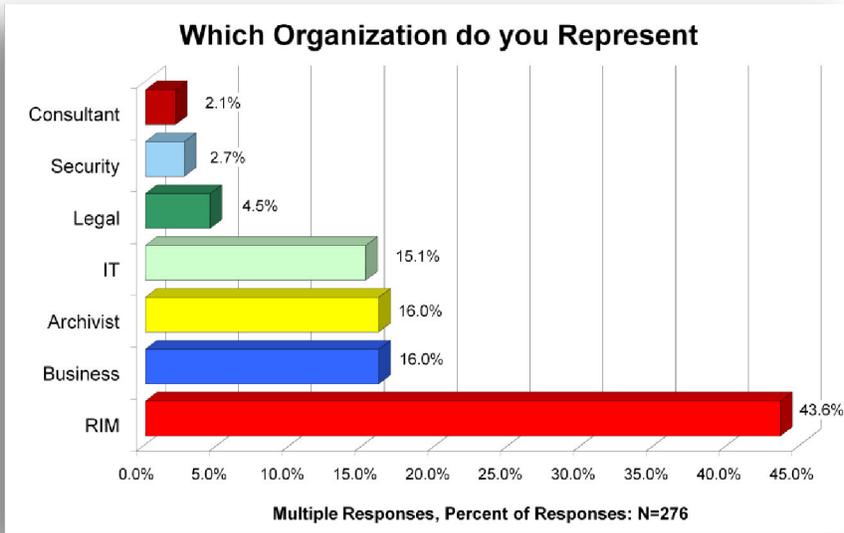
JOB TITLES: The mix of respondent job titles was very broad. The majority of respondents (40%) carried a RIM title. Adding balance to the survey, were a number respondents belonging to a business group with titles like application specialist, business analyst, researcher, and even seven (7)-'C-level' individuals. Overall, a broad mix of respondents participated, including some from legal and security.

Figure 12



OPERATING ORGANIZATION YOU REPRESENT: In contrast to job title, this listing of actual responsibilities allowed multiple responses per person since many people have jobs that overlap multiple disciplines. This is an important perspective, "What organizational groups are represented in the survey?" RIM, business, IT, and Archivists dominate. The understanding of responsibility becomes clearer in the next question, "What are your job responsibilities relative to long-term retention?"

Figure 13

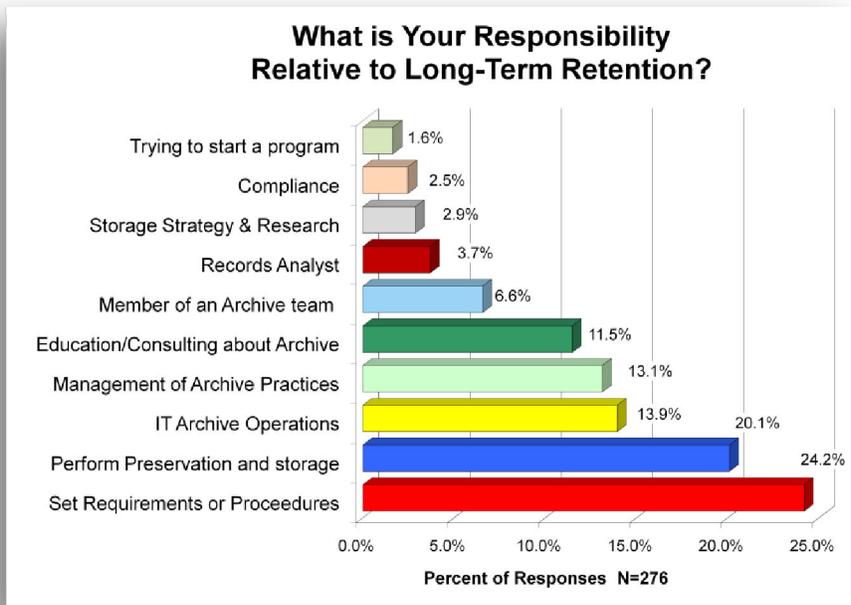


100 YrATF Analysis:¹⁰ Does this imbalance of RIM respondents introduce any identifiable bias in the results? Analysis of the data

by respondent organization showed differences by organization regarding factors such as their view of drivers and requirements. But, since the study is not seeking to focus on differences, the bias is irrelevant. Instead, the study seeks to aggregate the collective wisdom. All viewpoints are needed and in that respect the study was successful in getting many types of people to participate.

RESPONSIBILITY FOR LONG-TERM RETENTION: Respondents were asked what their responsibilities are in their organizations. The top four responses correspond directly to the population of RIMs, Archivists, and IT respondents in the survey.

Figure 14



An open-ended question was attached to this one asking respondents to explain their responsibilities. The interesting responses follow on the next two pages.

¹⁰ 100 Year Archive Task Force – this label is used to mark analysis & opinion as compared to prior discussion that is only reporting on survey results.

RESPONDENT COMMENTS: What are your responsibilities relative to long-term digital information retention?

I participate in setting retention policy and retention periods. I assist the IT department and records owners in establishing procedures for handling and retaining documents, to be sure their processes comply with our retention schedule.

Auditing of compliance with records retention policy.

I develop and implement policies and procedures governing the selection and preservation of all company records that have long-time archival value.

I recommend preferred media for long-term storage of records, based on the record type. There is a problem with the above question - long-term is anything over ten year retention.

Responsible for the management of all agency records, regardless of format - includes but not limited to storage, retrieval, preservation, compliance, format specifications and procedures, policies, authentication, etc.

Responsible for formulating policies and procedures for the enduring preservation of electronic records scheduled for archival retention for historical or other reasons.

In preserving digital design objects, we commit to preserving the contents of the files both functionally and at a bit level. Right now, our strategy is to implement a two-tiered collection: one that contains all of the native data (i.e., CAD formats, animations, renderings, etc.) that is monitored in DSpace for bit-level preservation via checksums and one that contains selected 'output' files that represent significant steps in the design process. These latter are translated into pdf or tiff formats and preserved on a functional basis as art objects.

As of right now, I am the sole person responsible for the translation/migration/preservation of these objects in the Architecture and Design Department; our IS department is responsible for maintaining the server on which these data are stored and the web interface for the collection management system.

I am required to acquire, preserve, catalog and make available information in all formats that in any way documents the history, people, and institutions of the West Florida region. These include publications regardless of format, many now emerging as CD/DVD, collecting records of agencies (families, churches, etc.) many of which may be digital in storage (databases, etc.). My agency has been in existence for 40 years.

**Respondent
Comments**

What are your responsibilities for long-term information retention?

Respondent
Comments

What are your responsibilities for long-term information retention?

I am responsible for maintaining the institutional archives and managing the records management program. Digital information responsibilities are shared with the IT Department and a network of professionals interested in evolving records management policy, procedures and best practices (representatives from IT, Communications, Legal and Audit). Establish policies that assist in making a long-term archive decision. Specifically, a risk analysis on whether a particular record will be required for greater than 1 year because of business, technical or compliance purposes. By establishing evaluation criteria that is tailored to our business, we hope to get most of our requirements documented and start looking for tools that can help us meet our practices. I keep current on standards, best practices, research etc. on digital preservation from the records management perspective and share that with decision makers to balance similar information from the IT perspective.

Storage Management, Disaster Recovery, Server Management, Backup & Recovery, & Business continuity Planning

(I am responsible for) delivering a Preservation Plan from analog to digital preservation of 35+ years of media recordings.

I am the archivist charged with the long-term retention of digital records. The mission of our organization is to preserve, protect and provide access to the documentary evidence of the history of our industry.

I help establish retention policies and manage the records transferred to the Archives for both short- and long-term preservation.

Develop Retention Periods, preservation and migration strategies and compliance procedures for our organization's information assets.

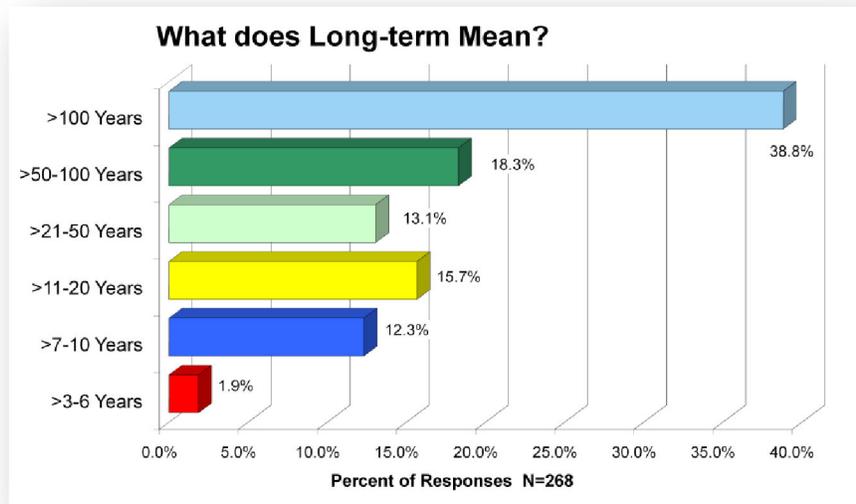
SURVEY RESULTS

This section of the report presents the summarized data from each of the questions asked on the survey along with analysis where appropriate. Interpretive analysis is marked as the “100 YrATF¹¹ Analysis”.

WHAT DOES LONG-TERM MEAN:

The survey question was intentionally open-ended and did not provide a definition of “long-term”. The intent was to capture respondent’s beliefs without creating a bias by predefined categories. However, the results in Figure 15 are biased because of the influence of a high percent of RIM and Archivist respondents. See the Figure 16 for an analysis by job function. Overall, 98% of respondents called long term over 7-10 years.

Figure 15



100 YrATF Analysis: It is hard to accept that “long-term “means over 7 years by strictly reading this data. The challenge is that a more meaningful definition is needed because it does not align with experience. In discussing this problem one-on-one with archive practitioners, their experience identifies that a threshold does exist beyond which retention becomes real hard. Retention periods of less than 10-15 years are usually considered achievable with today’s IT practices and periodic assisted migration. Beyond this timeframe, multiple migrations are required and the potential of losing or corrupting information increases rapidly. So, for the purposes of long-term digital information retention, the 100 YrATF is using the time period ‘over 10-15 years’ as the definition of long-term.

What does Long-Term Mean?

Over 10-15 years

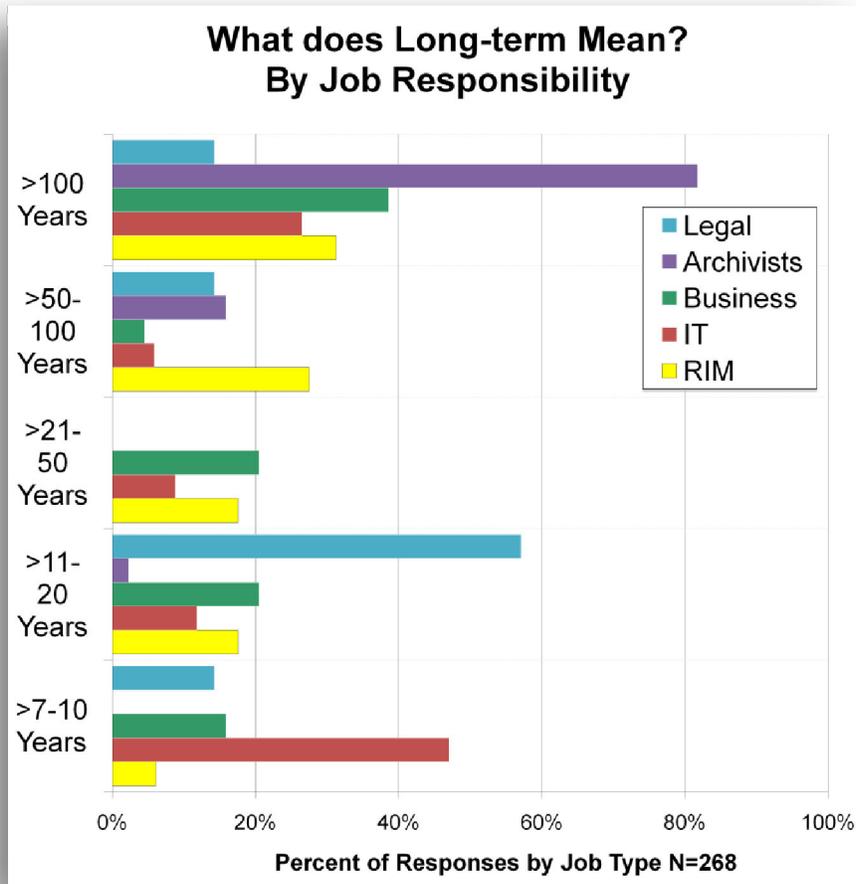
¹¹ 100 Year Archive Task Force (100 YrATF)

does not address the historic needs. It is a matter of preserving for future generations the important history and records of an organization. This is the problem with the 'Digital Archive', you are not thinking long enough into the future. (Source: Respondent)

Definition of long-term by job:

In Figure 16, the data from Figure 15 is analyzed by job responsibility. This data amplifies why collaboration is required since it is clear that each group has a different viewpoint. For example, RIM and Archivists are aligned - 60% of RIM and 97% of Archivists say long-term is more than 50 years. Whereas, 47% of IT says it is 7-10 years.

Figure 16



100 YrATF Analysis:

These responses demonstrate that the IT frame of reference is very different from RIM or Archivists. Whatever the reason, the discrepancy in results confirms the need for collaboration between these groups to understand and help solve their long-term retention problems.

EXTERNAL DRIVERS

The next series of questions looks at the external drivers for long-term retention and how each factor is defined. Data is analyzed in summary form, as rankings by organizational responsibility, and by asking for definitions of what each driver means. This comprehensive look at drivers for long-term retention is aimed at better understanding the organizational issues and their requirements.

External factors driving long-term retention

The external drivers for long-term information retention are important. They explain why requirements exist for many parameters such as retention periods, confidentiality, security, integrity, and protection, they identify business risks, and they provide justification for budget and resource allocation to address retention programs. The top five (5) drivers identified are business, legal, security, compliance and other risk (the ‘other-risk’ category is principally “the risk of losing an organization’s history”).

100 YrATF Analysis: The variances by job responsibility are interesting but of no significance other than to further confirm the importance of collaboration and capturing all perspectives.

Next, respondents ranked the importance of these drivers based on their organizational responsibility. The differences are interesting and further reinforce the need for collaboration. Table 3 shows that fear of losing the organization’s history is the top concern in the business group, compliance is the top concern for RIMs, and legal risk is top for IT, security, and legal. At the other end of the scale, security is not high in importance to anyone other than the security group. That merely means the other groups have larger issues such as legal risk.

Figure 17

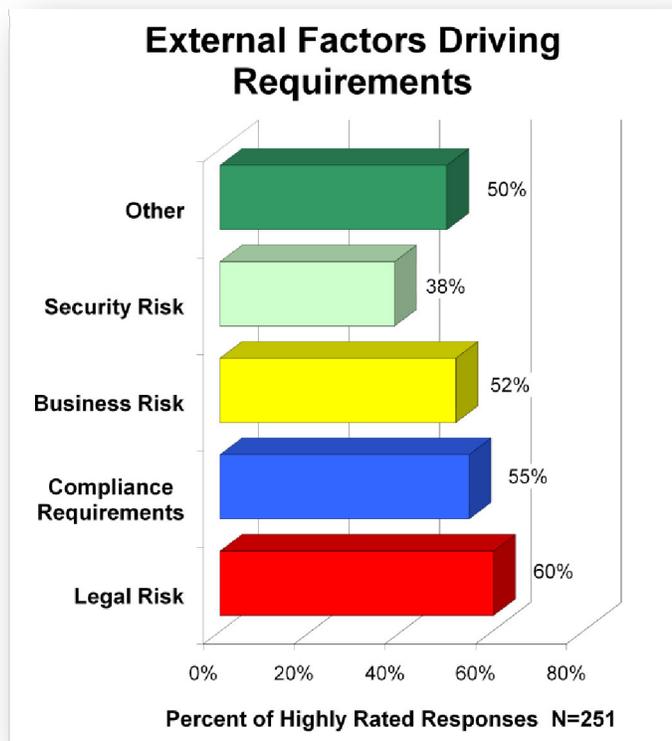


Table 3
View Of Top External Factors As Drivers For Long-Term Retention by Organization
 (Ranking 1-5, 1-High, 5-Low)

Organizational Responsibility	Business Risk	Legal Risk	Security Risk	Compliance Risk	Other* Risk
Business	2	3	4	5	1
IT	3	1	4	2	5
RIM	3	2	4	1	5
Legal	3	1	4	2	5
Security	3	1	2	4	5

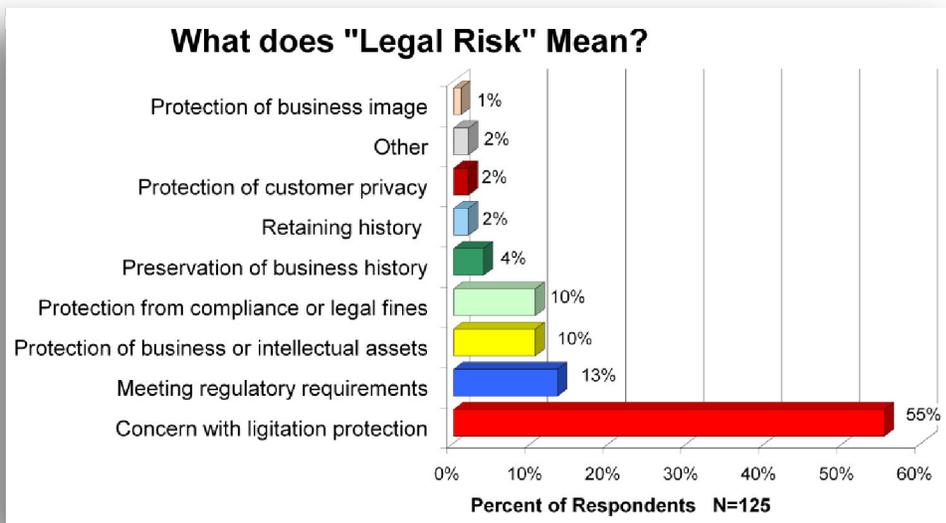
* Other means principally – risk of losing the history of the organization

The next series of questions break down what each driver means to the respondents. In these, differences by job responsibility show up again.

What does Legal Risk Mean?

Legal risk is principally considered to be associated with litigation and compliance costs including fees for non-compliance or non-conformance. In the definitions offered by respondents, concern with incurring fees, fines, or bad press from regulatory violations or legal judgments overwhelm all other issues.

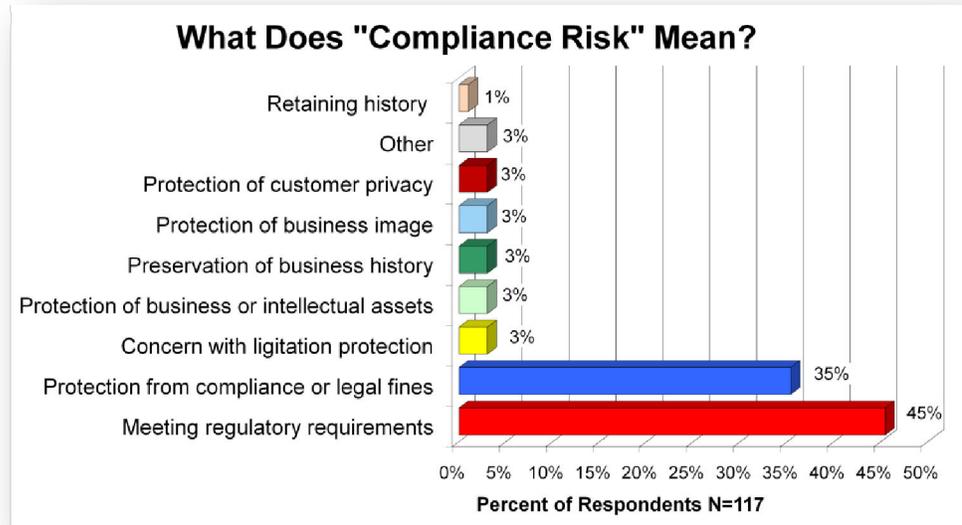
Figure 18



What does Compliance Risk Mean?

Compliance risk centers on fear of fines or loss of business reputation for non-compliance.

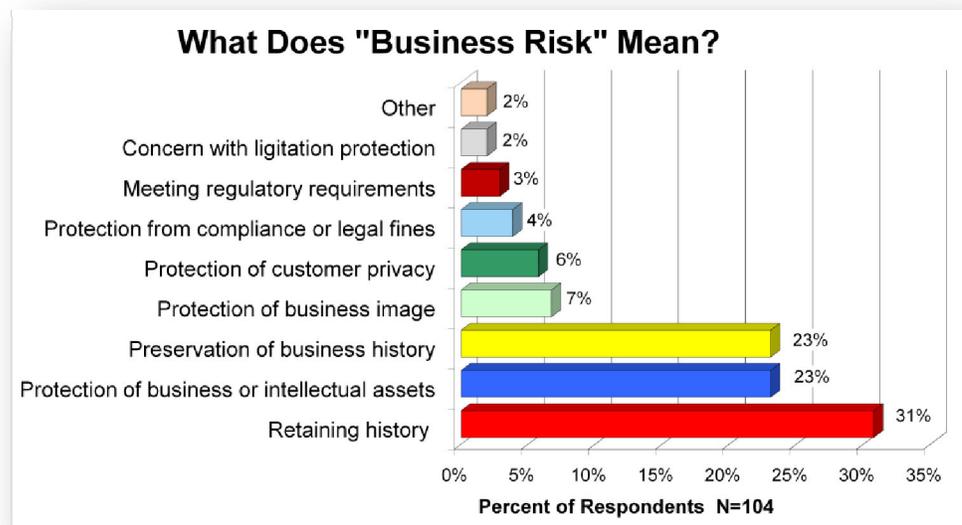
Figure 19



What does Business Risk Mean?

The top business risk reported was fear of loss of business history. This response is likely due to the high percentage of RIM and archivist respondents who have this responsibility.

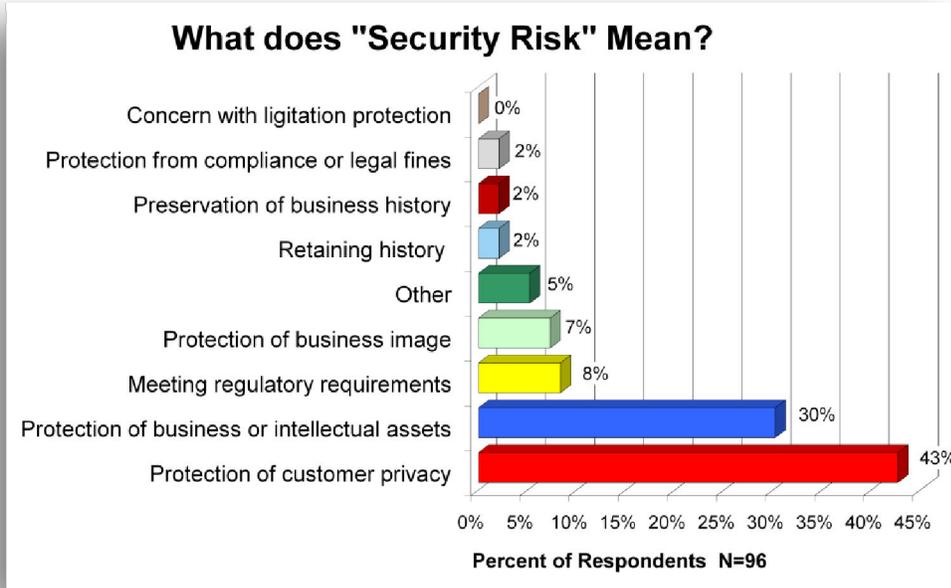
Figure 20



What does Security Risk Mean?

Security risk, in the context of long-term digital information retention, centers on customer privacy and the protection of business or intellectual property assets.

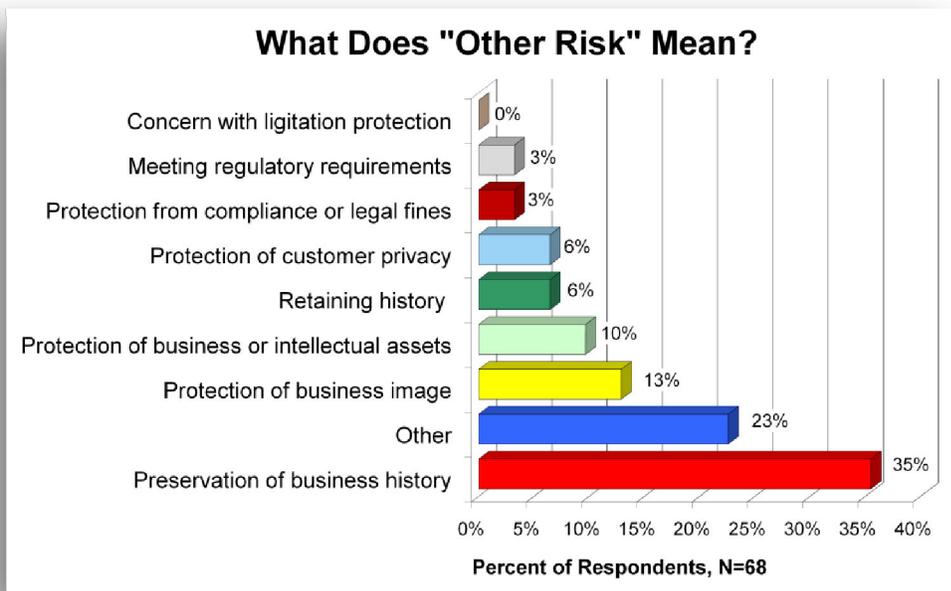
Figure 21



What are the 'Other' Driving Factors?

The responses for what comprises "Other risk" are principally the need to preserve an organizations history and its business or intellectual assets.

Figure 22



What are the drivers behind the need to retain digital information? RESPONDENT COMMENTS

The driving force for a true Archive is the preservation of the history of the organization for hundreds of years. Your survey does not address the historic needs. It is a matter of preserving for future generations the important facts of organization. This is the problem with the 'Digital Archive', you are not thinking long enough into the future.

The following factors are used to determine the retention of records: Administrative value to the organization, Operational value to the organization, Legal (contractual, statutory and regulatory), Financial, and Historical value to the organization. These factors are evaluated together in order to determine what is the required records retention period.

The rationale varies per record series. Are you using the IT definition of 'archive' or the RIM definition? Considerations include historical value, ethical considerations, liability protection, innovation, IP and other potential values.

We work mostly with public sector, government clients, therefore, preservation of public records for historical & legal purposes is the predominant driver but also the ability to re-use/re-purpose digital information as business assets for future delivery of products and services is a common driver.

Preserving historical documents (digital and non digital) for future generations and having copies of documents in digital form for access.

The same thing as drives any organizational archive--the need to select, manage, protect and make available over time the records of the organization needed for historical, legal, and/or administrative requirements.

The National Archives preserves the history of the actions of the Federal Government and protects the rights of our citizens.

Operational efficiency - time taken to search and retrieve information assets in digital form Historical value of the data that researchers and the general public will want accessible in the future.

**Respondent
Comments**

Long-Term Digital
Information Retention
Drivers

Top External Factors Driving Retention Requirements:

This chart provides a compilation of the factors behind the drivers and looks for commonality. For example, concern about meeting ‘regulatory requirements’ overlaps both compliance and legal risk and was the second most common issue behind concern about retaining the history of the organization for competitive and preservation purposes. These are the top drivers causing focus on implementation of adequate long-term retention practices. The top five factors behind all the drivers are:

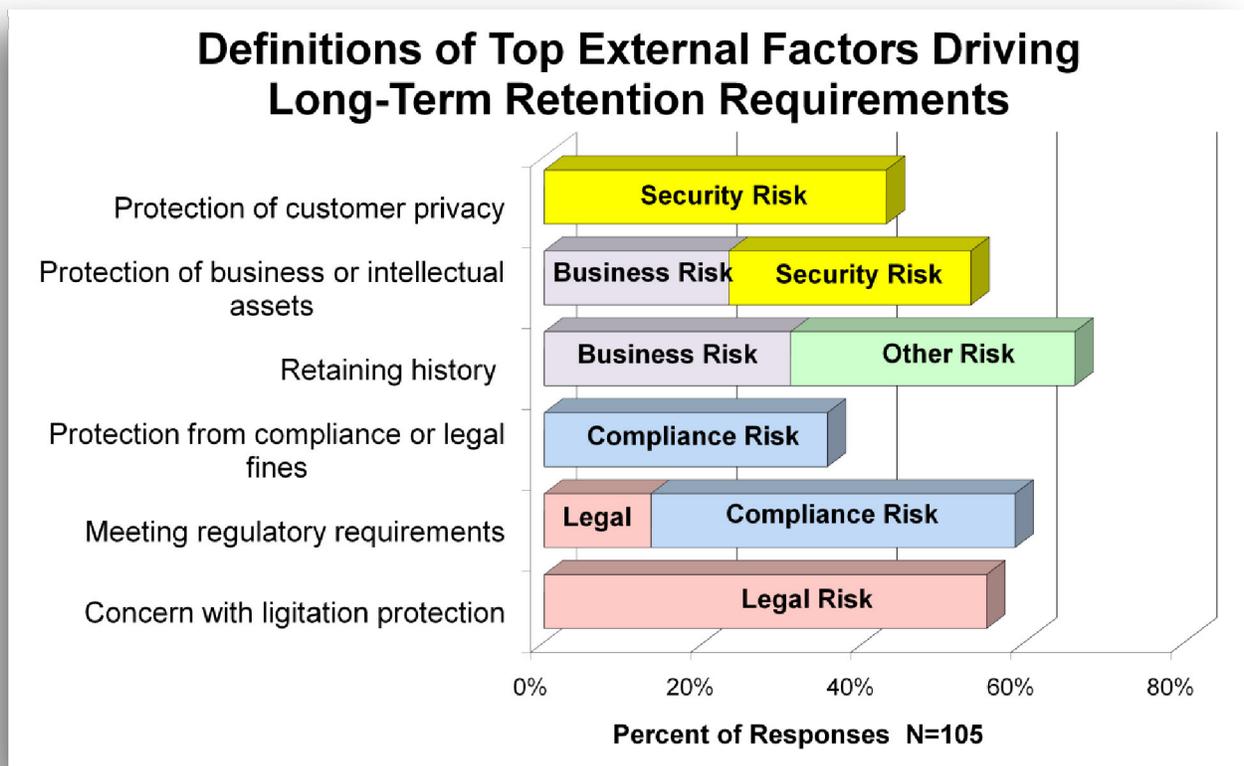
Retention Drivers

Solutions must meet the needs of all of them

- Protection and preservation of the organization’s history
- Meeting regulatory requirements
- Concern with litigation protection
- Protection of business or intellectual property assets
- Protection of customer privacy

100 YrATF Analysis: Any technologies or best practices being proposed as solutions to long-term digital information retention problems must also satisfy the needs of all of the business drivers.

Figure 23



INFORMATION PROFILE

The next section of the survey focuses on discovering the information profile of the respondent's sites including how information is managed and retained.

Who Defines Requirements for Business and Compliance Information

These charts show some interesting answers to the question of who defines requirements for business and compliance information.

14% said "All" which is a good indication that collaborative efforts are in place at some organizations. The top organizational groups defining requirements are the business groups and RIMs who define requirements in over 40% of the respondent's organizations. Legal only defines requirements for 15%-18% of the respondents. It is often assumed that legal 'controls' the entire requirement setting process, but that was not the case with this set of respondents. 14% said IT defined requirements. This is a low result, but consistent with other data in the survey such as the analysis on Page 41 & 42 that explores how well organizations are structured to meet long-term retention requirements.

Figure 24

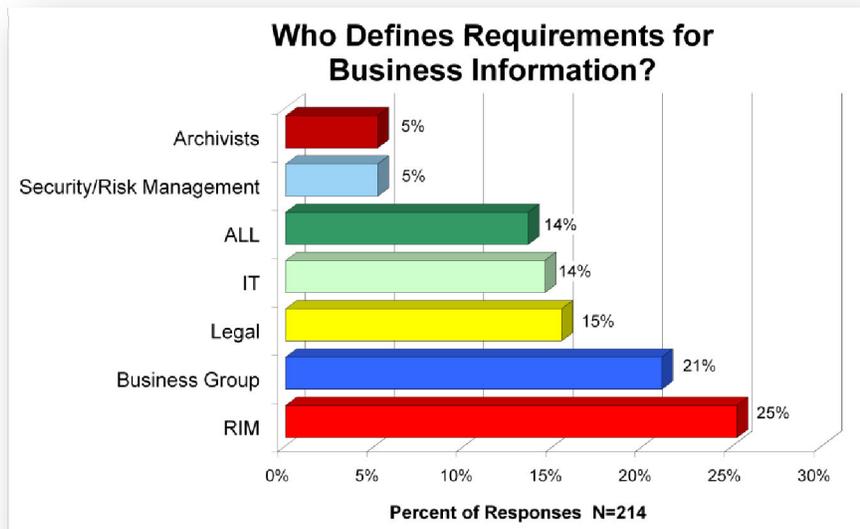
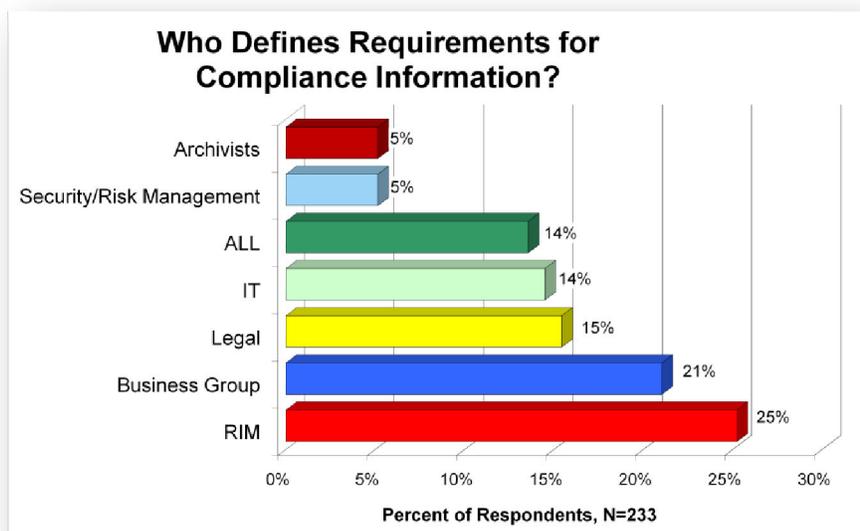


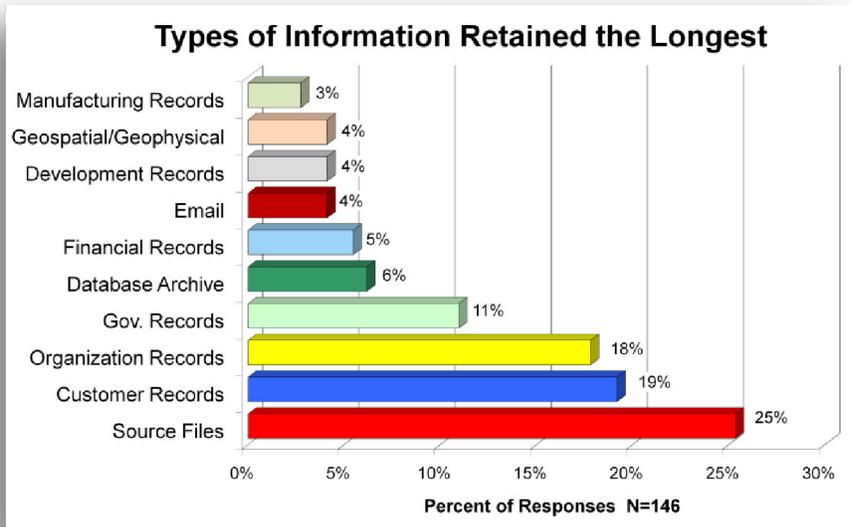
Figure 25



Types of Information Retained the Longest:

Source files, customer records, and the organization’s records comprise 62% of the responses for the type of information being retained the longest. Few said information such as email or manufacturing records are kept the longest. “Source files” is an archivist term relating to the ‘original’ files or ‘information objects’ not backup, redundant copies, or unverified duplicates.

Figure 26



Each of the information types that are retained the longest have corresponding external drivers. For example, source files are controlled by compliance, legal, and business interests. Customer records have compliance and business requirements. And, organizational and governmental records have both compliance and historical value.

100 YrATF Analysis:

These results are highly dependent upon the profile of the respondents and should be used carefully. Generally speaking, the types of records that have the longest retention requirements are relative based on the type of organization and the specific compliance rules governing its business. Archivists and RIMs are most concerned with ‘source files’, the originals. IT would be more focused on databases, financials, or customer records. With the large percentage of RIM and Archivist respondents, it is not a surprise to see “source files” as the longest retained information type. What is a surprise in the data are the 6% who put ‘database archive’ records on top. This aligns with the data in Figure 29 on page 34. Structured information can not be overlooked as a key information type in the requirements for long-term digital information retention.

Longest Retention Requirements

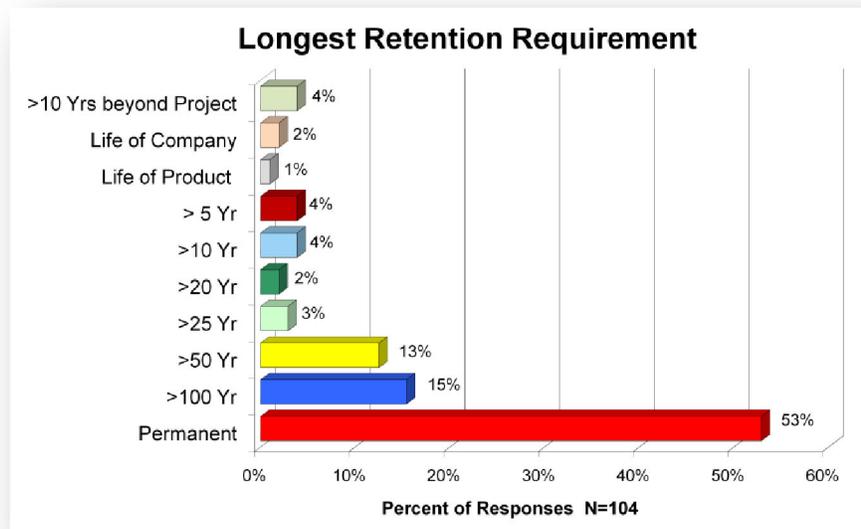
The survey next asked, what is your single longest retention requirement? The results are impressive. 53% of the respondents said they have information that must be retained permanently and 83% said over 50 years.

The long-term retention needs are real.

Figure 27

100 YrATF Analysis:

The need for long-term retention is greater than expected. These results also point out the need for classification and collaboration with IT as different information types have widely different retention requirements.



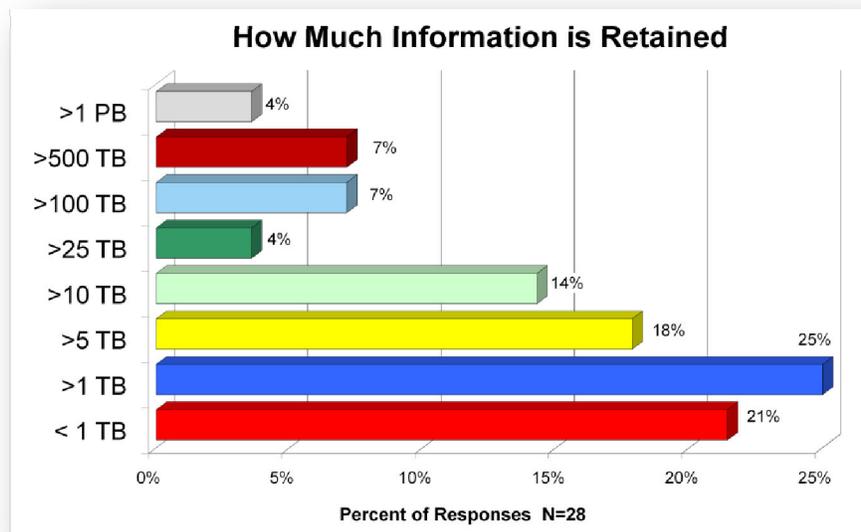
How Much Information Is Retained?

Most archives at these respondent's sites are small, less than 5 TB. However, 18% said their problem is over 100TB. The 18% are validation that large-size, data center-based repositories are in operation. Even a couple petabyte size repositories participated in the study.

Figure 28

100 YrATF Analysis:

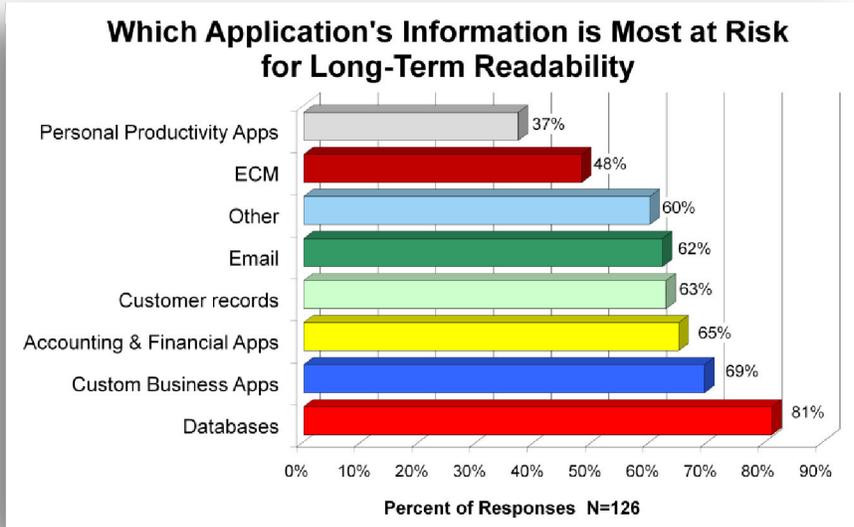
As said in the slide, the high percentage of archivists in the study biases the size of the information repository down. On one hand, 5 TB is small. On the other, it has all the same problems as a large repository – just smaller in cost and scale.



Which application’s information is most at risk for long-term readability?

Here is another important message in the survey data – all applications are at high risk. Database information is most at risk according to 81% of the respondents. The next three classes of information (custom, financial, and customer records) are usually

Figure 29



also built on databases as well. If the data were recompiled in this manner, then email and document management risk would be well below database risk.

100 YrATF Analysis:

The industry’s focus on retaining unstructured data and email has left an important gap. According to these respondents, databases (including all enterprise applications running on databases) are at far more risk than any

other type of information. The second point in the data is that all applications have long-term retention risk. Nothing is safe, providing more validation that this problem needs to be solved.

Long-Term e-Mail Retention Practices

The retention practices for e-Mail were looked at from two perspectives. First, are there different types or classes of e-Mail from a retention perspective and second, what percent of organizations retain e-Mail records over 10 years. This data confirms the importance of classification as requirements for e-mail vary significantly and correspond to the degree of regulation the organization operates under.

Figure 30

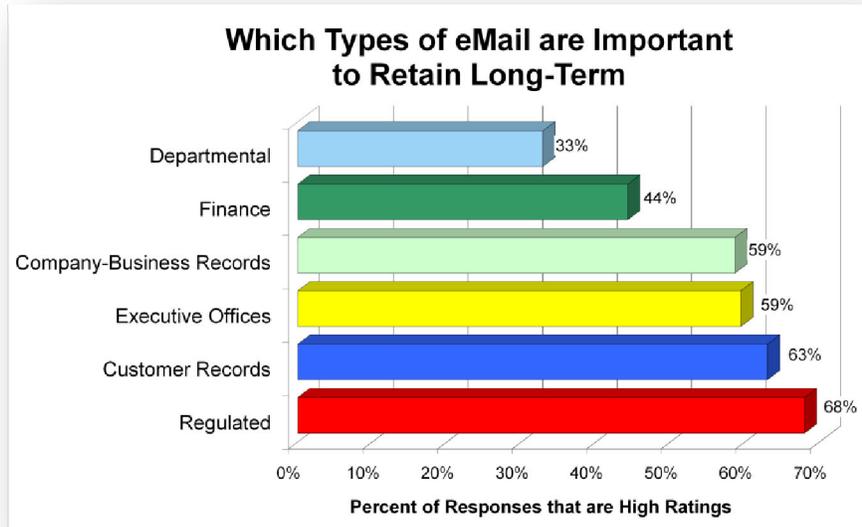
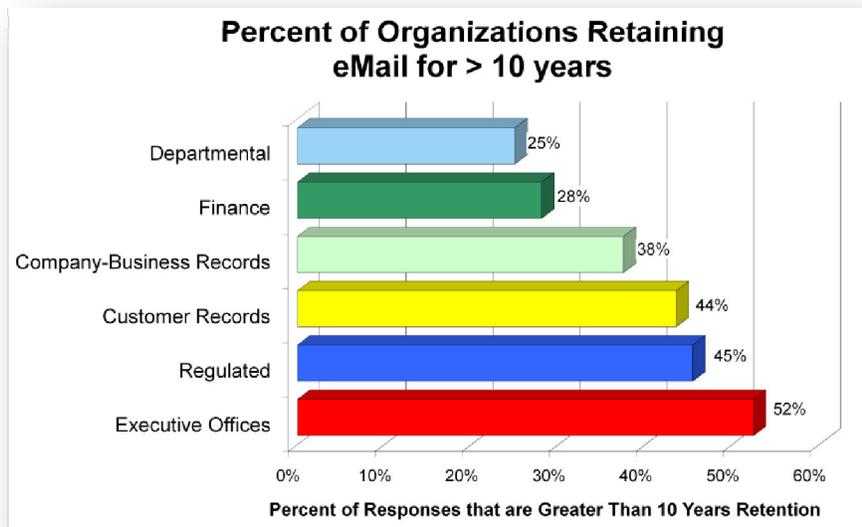


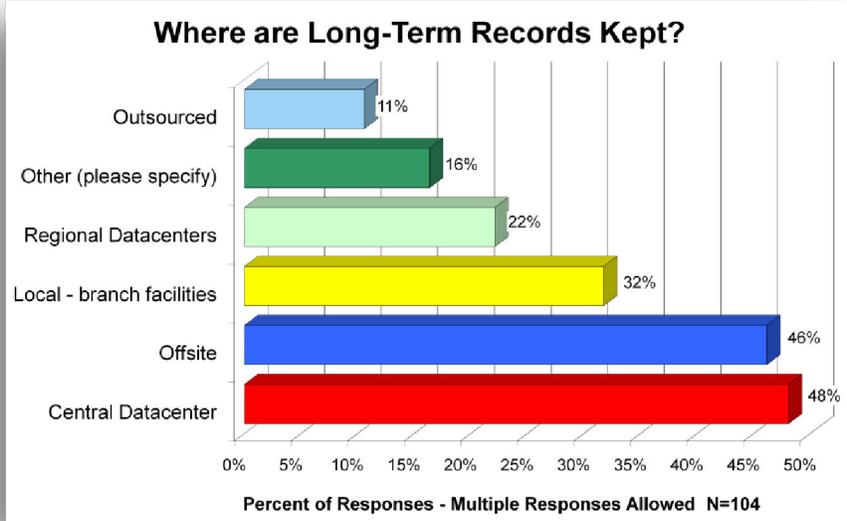
Figure 31



STORAGE INFRASTRUCTURE

The next section of the survey focuses on the storage infrastructure for long-term digital information retention.

Figure 32



Where Are Long-Term Records Kept?

Central data centers and offsite disaster recovery sites dominate the location of long-term repositories. The 'local-branch facilities' response from the RIM and Archivist may be a terminology issue. IT professionals would categorize most of the 'local' sites as 'central' or 'regional' locations.

What is the Retention Period For Data Stored on Tape

The use of tape in long-term archives is legendary. 37% said it was their permanent storage media. Yet, many respondents were negative towards tape. In addition, their reported migration practices were very inconsistent.

Figure 33

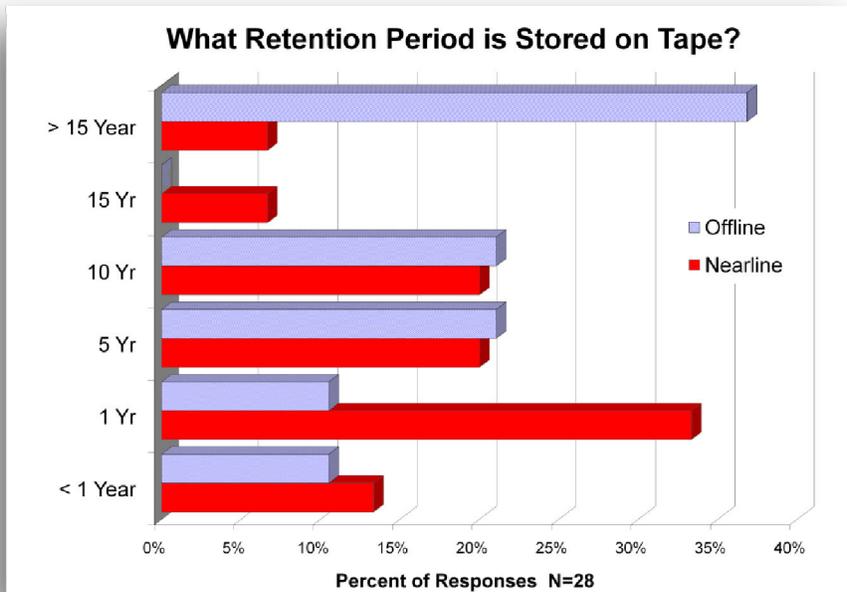


Figure 33 compares offline tape storage to nearline tape storage and shows very different use models for long-term retention. For example, information stored nearline where it is accessible is retained for far shorter periods than information stored offline. Why? 37% of respondents consider offline tape as their permanent archive media. In comparison, 46% of respondents keep information on nearline tape for one year or less.

TAPE PHYSICAL MIGRATION: Respondents were asked what physical migration cycle is used? 21% of the respondents said they do not migrate at all and 11% said they are moving away from tape entirely for long-term retention.

Table 4
**How Do You Physically Migrate Data On Nearline Tape
 Used For Long-Term Retention?**
 Percentage of Responses, N=28

Unknown	39%
We don't Migrate	23%
Copy onto compatible format	14%
Moving away from tape	11%
Other	14%

100 YrATF Analysis: One thing this means is that tape is used in various ways in the storage tier as best practices seem to be very inconsistent. For 21% of the respondents to say they do not do physical migration, puts a lot of information at risk.

RESPONDENT COMMENTS

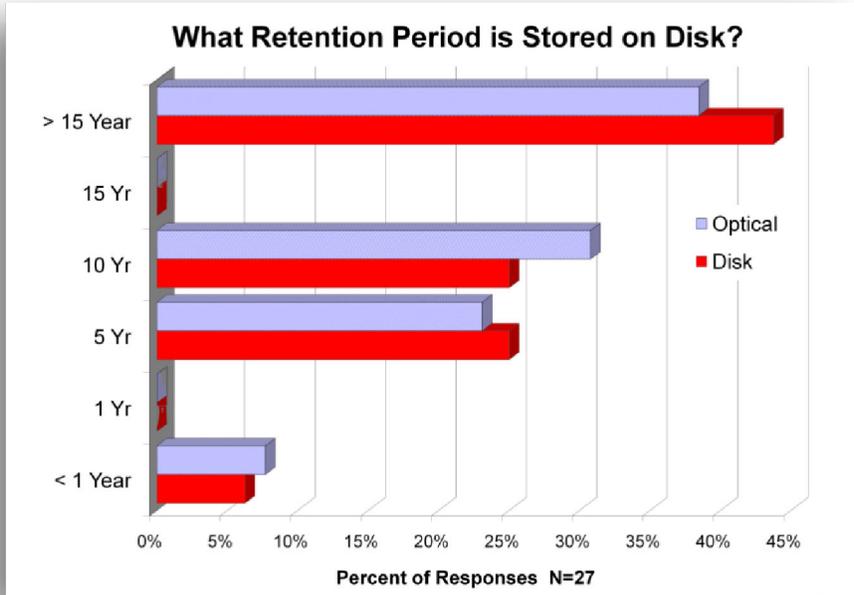
- We do not use tape for long-term storage.*
- There must be a high level of error checking to insure no information is destroyed or changed.*
- We are moving to disk storage and moving away from offline. This is an ongoing initiative.*
- Tape is for backup purposes only. It does not provide an appropriate means of storing records.*
- We don't migrate. We don't keep information long enough to migrate because I recommend against storing long enough to need migrating.*

**Respondent
Comments**

Migration on Tape

What is the Retention Period For Data Stored on Disk Media?

Figure 34



The retention profile of optical disk and hard-disk is very similar with one interesting surprise; more people said they have greater than 15-year retention on hard-disk than on optical media.

Approximately 40% of disk and optical disk responses are using these technologies for retention periods greater than 15 years. Migration practices are much more in line with the National Archive & Records Administration (NARA) requirements¹² than were the tape responses. But, 22% still claim that they do not ‘migrate’. 22% is a very big percentage.

DISK PHYSICAL MIGRATION: When asked what physical migration cycle is used, these results were given.

Table 5
How Often Do You Migrate Data Retained On Disk?
 Percent of Responses, N=27

Unknown	41%
3-5 years	30%
Continuously	7%
When Systems are Upgraded	4%
We don't Migrate	22%

¹² NARA requires migration on disk, every 3 years and on tape, every 5 years.

RESPONDENT COMMENTS

Once on CAS there is no migration except to create a replica on another CAS at a remote site.

Records should be retained based on their content, not on their format (or storage method, in this case). A migration plan is required for all records stored in an electronic information system that has a retention period greater than 5 years. Refreshment and migration strategies are established based on the nature of the records and the storage method employed. Usually, refreshment every three years and migration every five.

(Migrate) When technology obsolesces

**Respondent
Comments**

on Disk Migration

OPERATIONS SATISFACTION STUDY

The next section of the survey analyzes satisfaction or dissatisfaction towards the operating areas of cost, migration, security, discovery, and organizational structure. How satisfied respondents are with their organization’s internal capabilities tells an important part of the total story and helps identify needs.

Note: The format of the next set of charts is based on the percent of responses that rated their satisfaction or the importance of their programs either high (in agreement) or low (in disagreement) on ‘Likert-scale’ ratings. (Rate 1-5, 1 low, 5 high) The best way to interpret each chart is to consider the percent high against the percent low. To interpret preferences or opinions look at each end of the scale to evaluate results. For example, a response that is 50% high and 20% low indicates that the most respondents are in strong agreement. The missing neutral responses are essentially statements of ‘I don’t know’ or ‘I don’t feel strongly one way or the other’.

What is Your Satisfaction with Programs your Organization is Doing to Reduce The Cost of Long-Term Retention?

Figure 35

Programs to Reduce Cost of Retention	Percent of Responses that are High Ratings	Percent of Responses that are Low Ratings
Classifying Information's retention periods	58%	18%
Only keeping "important" information long-term	38%	32%
Coordinating the business IT RIM & security to classify info & set requirements	33%	30%
Manually Purging expired data	32%	28%
Classifying Information on creation	31%	43%
Complaining about cost	28%	43%
Setting rules for automatically purging expired data	25%	48%
The business IT RIM & security are working to solve the cost problems	24%	33%
Moving to dedicated archival systems	21%	46%
Nothing	19%	61%
Leaving it to IT to figure out	19%	55%
Implementing charge-back for long-term retention	7%	78%

N=129

Several important points stand out in Figure 35. At the ‘this is very important’ end of the ratings, classification, collaborative efforts, and eliminating expired data jump out as the focus area for reducing cost. At the ‘not important’ end of the ratings, ‘charge

back’ is not a preferred method. And, 20% were in agreement that they are doing ‘nothing’ or that it is IT’s problem.

What is Your Satisfaction with Your Organization’s Ability to Solve Media Migration and Long-term Readability?

Very little more needs to be said about migration other than the responses to these questions. These responses validate that migration is a huge problem that is not getting a lot of attention outside of people casting around looking for better approaches.

Figure 36

Programs to Solve Migration & Readability	Percent of Responses that are High Ratings	Percent of Responses that are Low Ratings
Migration is a huge problem and we need help	49%	28%
Looking for better systems that will handle migration	36%	35%
Overwhelmed with capacity growth in the archives	31%	40%
Unable to cope with the load and cost of migration	27%	44%
Pretending it will go away or it is someone else’s problem	24%	62%
Hoping IT has it figured out	22%	56%
Writing images and in an archive format with embedded metadata (XML PDF-A)	22%	52%
Spending a lot of time and money	21%	50%
Archiving systems and applications to assure we can read the information	20%	50%
Nothing	17%	65%
Implemented pre-planned media migration schedules	16%	63%
Following the OAIS standard	11%	66%

N=125

100 YrATF Analysis: The messages in this chart are very important. ~50% are in agreement that migration is a huge problem and they are not spending time and money on fixing it. ~20% agreed that they are doing nothing to solve the problems other than hoping someone else has it figured out. 27% said that they agreed that they are unable to cope with the load and cost of keeping up with migration. Only 22% said they are trying to write long-term archive formats. What are the rest doing? Not enough to assure survivability of their information long-term.

How Satisfied are You with Your Organization’s Security & Privacy Controls to Reduce Long-Term Retention Risk?

The responses indicate that encryption and classification are well established in the practice of secure long-term retention. Only 16% agree that they are doing nothing.

100 YrATF Analysis: These results are not what are expected from the IT community’s perspective. The apparent inconsistency in this data may be that RIM and Security people generally define and operate security controls and IT does not.

Figure 37

Programs to Reduce Risk	Percent of Responses that are High Ratings	Percent of Responses that are Low Ratings
Handling 'highly confidential' information separately from 'non-confidential'	46%	29%
Coordinating the business IT RIM & security to classify info & set requirements	43%	29%
Controlling access to confidential data via separate authentication practices	36%	37%
Classifying Information based on Security & Privacy on creation	36%	36%
Re-Classifying Information when its value changes	30%	44%
Encrypting all confidential data in transit	27%	52%
Handling all information the same	22%	59%
Encrypting all data in transit	16%	62%
Nothing	16%	68%
Encrypting all confidential data at rest	15%	65%
Loosing track of or control of growing numbers of encryption keys	12%	73%

N=123

How Satisfied are You with Your Organization’s Legal Discovery Capabilities?

eDiscovery is a big issue according to 37% of the respondents and on the inverse side, 30% said that they do not have eDiscovery programs in place. ~33% of respondents said eDiscovery is a big challenge in the long-term digital information repositories they manage. And, 19% are deploying special purpose repositories that have discovery capabilities integrated.

Figure 38

Satisfaction with Legal Discovery Programs	Percent of Responses that are High Ratings	Percent of Responses that are Low Ratings
Struggling to search the long-term digital repositories	37%	38%
Struggle to find all the distributed archival media	34%	39%
No Discovery Programs	29%	62%
Working to automate discovery across the long-term archive repositories	25%	53%
Always able to find all information requested within the allotted timeframe	23%	51%
Deploying special purpose digital archives with discovery capabilities	19%	62%

N=116

How Well is Your Organization Structured to Meet Long-Term Retention Requirements?

This question probed the existence of collaboration and the role security and legal professionals play in setting requirements.

- 20% agree that IT is autonomous
- 49% agree that security has an important role
- 15% agree that legal is in charge – this corresponds to the results presented on Page 31

Figure 39

Satisfaction with Organizational Structure	Percent of Responses that are in High Agreement	Percent of Responses that are not in Agreement
Security has an important role in setting requirements	49%	14%
IT is closely Aligned with the Business Group	44%	17%
IT is closely Aligned and coordinates with RIM	35%	31%
IT, RIM, security, and the business never talk about requirements	23%	57%
IT is autonomous and has to guess or make up its own rules	20%	50%
Legal runs the show	15%	43%

N=121

100 YrATF Analysis: This data indicates that many companies have not organized or established collaborative responsibilities to manage their long-term retention and preservation requirements.

Are You Satisfied With Your Organization’s Long-Term Retention Methods?

This set of questions queried respondent’s satisfaction with their long-term retention practices.

Figure 40

Satisfaction w/ Long-Term Retention	BUSINESS	IT	RIM
Retaining value to the organization in its long-term archives	47%	51%	41%
Storage systems used for long-term retention	34%	53%	25%
How well the organization works together to assure retention and readability	20%	49%	22%
Ability to access & read information in 7-10 years	33%	48%	24%
The cost and ability to migrate data to newer media technologies	24%	35%	18%
Ability to comply with legal discovery requirements across the various repositories	20%	30%	20%
The cost to maintain long-term archives/repositories	18%	21%	26%
Ability to access and read information in 50+ years	19%	21%	17%

N=53

100 YrATF Analysis The results corroborate the message received many times that improvement is needed on many fronts.

- IT is far more satisfied than RIM or the Business Group. This could be because they define “long-term” as a shorter period than RIM or the business does.
- Business and RIM are very dissatisfied with the collaboration with IT to assure long-term retention and readability.
- Very few IT respondents are satisfied that they can access and read information over the long-term.
- No one is happy with the cost of maintaining long-term information access. It is important to remember that responsibility for information risk lies with the business and not IT. Only the business group understands the value of the information.

Are You Dissatisfied With Your Organization’s Long-Term Retention Methods?

Figure 41 tabulates data that is the inverse of the previous question. “How ‘dissatisfied’ are you?” In the previous question, ~50% claimed satisfaction with their ability to retain value in the long-term archive and 50% of IT claimed satisfaction with their storage systems. Here less than 20% of IT state they are dissatisfied compared to 43% of RIMs. When asked specifically about the ability to access and read information in 50 years, 60% are dissatisfied.

100 YrATF Analysis: What is different between IT and RIM responses? It is known that they have different expectations and experience. IT thinks of long-term as greater than 7 years and RIM as greater than 50. In addition, most IT respondents do not know the business requirements for retention, while the RIM community is focused on information value and retention and have a clearer awareness of what ‘successful retention’ means.

Figure 41

Dissatisfaction with Retention	BUSINESS	IT	RIM
Retaining value to the organization in its long-term archives	26%	19%	43%
Storage systems used for long-term retention	34%	16%	57%
How well the organization works together to assure retention and readability	50%	23%	59%
Ability to access & read information in 7-10 years	40%	28%	50%
The cost and ability to migrate data to newer media technologies	42%	33%	45%
Ability to comply with legal discovery requirements across the various repositories	48%	35%	61%
The cost to maintain long-term archives/repositories	53%	32%	37%
Ability to access and read information in 50+ years	57%	55%	68%

N=53

The most surprising data in this chart is that only 55% of IT (compared to 68% of RIMs) are dissatisfied with their ability to access and read information in 50 years. Perhaps, this is confirmation that IT really doesn’t understand the problems associated with long-term retention.

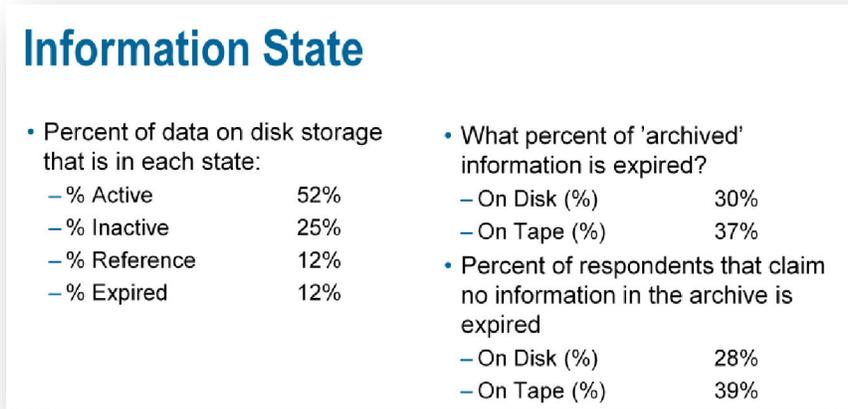
What is Your Profile of Information State?

The survey asked: respondents to profile their current “information state” in their data center on server-disk storage (not including desktops). State was defined to include four use-modes: “active, inactive, reference, or expired”.

Figure 42

100 YrATF Analysis:

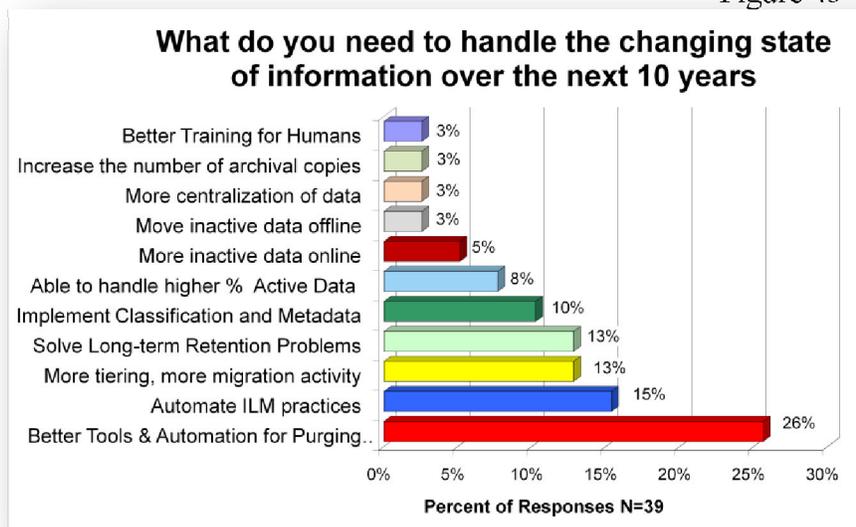
Unfortunately, RIM and archivists have different definitions for these terms than does IT which may have biased the results. Even so, the results are consistent with previous studies whose data ranges from 52% to 55% active, 20% to 25% inactive, 10% to 15% reference, 10% to 15% expired. Without robust eDiscovery, permanent deletion tools, and supporting practices, it is reasonable to see that 30%+ of ‘archived’ information was reported as ‘expired’. The paradoxical responses that claim that archives by definition do not contain expired information raise an important note. Either these people are doing a very good job of deleting expired information before archiving or their view is that everything in the archive is supposed to be there and will be held forever. This does not seem realistic for the data center and is actually not a good practice.



What is Needed to Cope With the Changing State of Information Over the Next Ten Years?

This question was asked to understand if the respondents had any expectations around changes to expect in the future. The responses indicate clear awareness of the growing magnitude of the management problems in dealing with the ever

Figure 43



increasing amount of digital information that has to be retained long-term and accessible. Answers included recognition of better tools, automation, classification, migration, training, practices, and data movement methods. They also include a small number of responses calling for better methods of handling a higher percent of active data. The responses are very insightful and supportive of the directions SNIA is taking ILM-based practices.

Respondent
Comments

What is Need to Cope
With the Changing
State of Information?

RESPONDENT COMMENTS:

Higher percentage 'active' (in the future). If not, (we will face) increased difficulty in being compliant with managing business records as a corporate asset.

(An) update process for deleting expired data. Data currently does not have retention period(s) added (as meta-data). Therefore, increasing the amounts of outdated data that may require discovery funding and may hinder litigation.

Need something that is easy to update to be readable on a long-term basis.

As we move deeper into a KM (knowledge-management) world, I expect that the next 10 years will see the differences between active, inactive, reference or expired' to change to useful or useless and be treated accordingly.

No need for change if effective, controlled procedures (are) in place.

Ensure digital data remains retrievable and readable. Our policy requires data to be tested on both criteria and to be printed in the event of evidence of deterioration.

The distribution of state on disk must match the ongoing business value of the data - automatically - if not; it's an unsolvable problem, since the humans cannot keep up with the data onslaught.

Active use of disk will hit same problem as tape - bits will require extra management in order to ensure integrity. Same issues of indexing, labeling and losing data because of information (not data) association will be lost over time. Need standards for higher order management functions.

Percentages need to better align between active and inactive with more expired data purged.

Need to be flexible to accommodate changes in regulatory requirements and business processes.

There needs to be a solution to the ever changing formats and versions which are non-compatible. This solution must be non-proprietary and insure no lose of data over the lifecycle of the records.

The volume of active and high-frequency accessed reference data will continue to grow rapidly - need better turn-around times on retrieval, especially from tape library storage, and increased capacity. If this doesn't happen, we will require acquisition and maintenance of multitudes of tape arrays/libraries, multiple SANS, etc., which will increase complexity in overall data warehouse infrastructure, and reduce usefulness of archived data.

Expired (information) needs to be purged otherwise unnecessary storage cost, legal exposure, and information overload.

Not as concerned with how it needs to change as in ensuring it occurs (does change). Uniform methods need to be developed and deployed for managing information based on its business and organizational value rather than its age and or perceived state. The state of information can change many times during its required retention, depending on the type and nature of work being done at any given time.

Fully automated is one answer. What happens if it doesn't change is that we will have to manage it better, ourselves.

Retain disk for active storage, migrate inactive content to lower cost system, migrate archival information to microfilm and to archival storage.

Better management tools for 'active' and 'less active' repositories.

The state of storage must be flexible and scalable, and can be migrated as the disks become obsolete.

There's no such thing as expired information if things are being managed properly, so that wouldn't apply.

We need proper lifecycle management of our digital records which is coordinated with the records retention schedule. Purge rules need to be implemented at the design stage of systems. Records of long-term value need to be migrated to alternate storage media which ensure accessibility over time (COM is still acceptable). The destruction (purge) of electronic records needs to be coordinated and authorized rather than being automatic.

Respondent
Comments

What is Need to Cope
with the Changing
State of Information?

Respondent
Comments

What is Need to Cope
With the Changing
State of Information?

We want to maintain up to 6 copies of archival versions in multiple locations through partners and possibly 1 offline copy. We want to keep 1-2 copies of reference copies.

From a records perspective, I need to maintain the recommended practice that disk storage not exceed 5 years. IT will want more inactive disk storage.

Systems that 1/ include disk and optical 2/ have intelligence based on rules that are specific to BU and through GED to individuals.

Data distribution will change more to inactive, reference. If expired data will not be handled, the volume will grow in the other areas 30 % per year.

We will be implementing an electronic records retention schedule once the document management system has been rolled out.

Better retention practices, including semi-active and disposition procedures.

IT needs to maintain the information and the metadata.

(Keep) long-term storage based on doc type/business application - not co-mingled.

(We are) looking to manage archived data dynamically, with software identifying expired data due for destruction.

More central coordination - fewer silos (of information or storage).

Inactive data MUST move to archive storage to avoid exponential growth in primary storage, and its associated costs.

5 years online 5-10 years nearline >10 off-line If this is not applied, (the) system will be full of unnecessary information.

REQUIREMENTS FOR LONG-TERM RETENTION

The last section of the survey covers respondent’s views of what the requirements need to be from several different perspectives and closes with respondents making recommendations to their peers.

What are the Top Requirements for Long-Term Retention?

This question gets asked in several different ways across the course of the survey to look at the question from different perspectives

Figure 44

and for consistency. Here are two views. First, a rating of business requirements in which long-term readability and accessibility rank highest followed by privacy, migration, and discovery. Now look at the table of “Top Pain Points”. The pain-point perspective illustrates a different way to look at requirements, yet produces similar results. Technology problems and operational problems lead the list. Business requirements don’t seem to make the top of the ‘pain’ list.

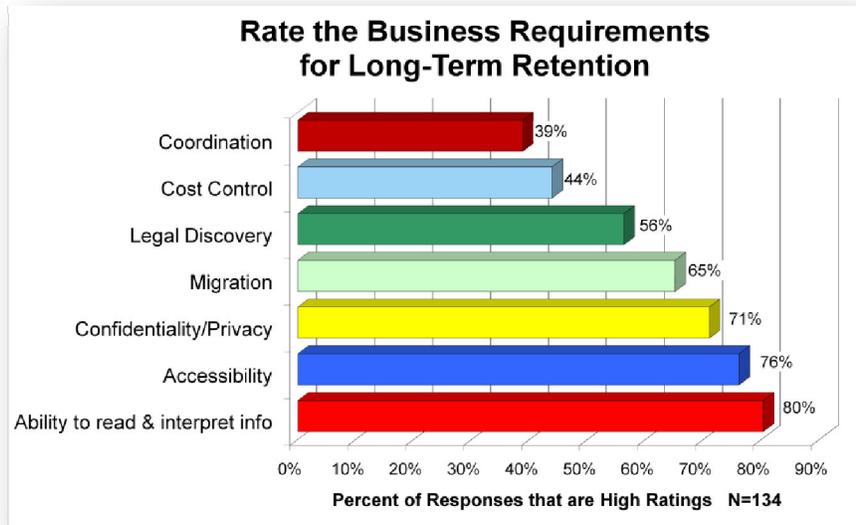


Table 6
Top Pain Points:
Percent of Responses N=144

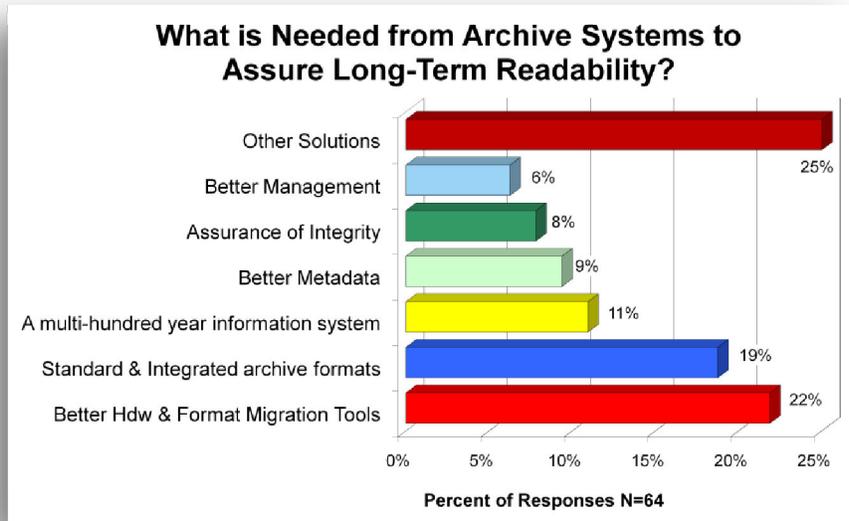
Media Migration	12%
Maintain Readability	10%
Technology Obsolescence/Upgrades	10%
Lack of Business Support/Commitment	10%
Cost	9%
Classification	8%
Lack of Collaboration	7%
Discovery & Deletion Difficult	6%
Too Many/Legacy Formats	6%
Lack of Expertise/Discipline	5%

100 YrATF Analysis: Between the top pain points and the top business requirements, the 100 YrATF now has a prioritized list of important requirements for any proposed solutions to the long-term retention problems. The top pain points provide excellent

validation of the requirements for long-term digital information retention. It is clear from this list that practitioners understand the problems and need help with solutions. Solving the technological problems of logical and physical migration are top on the list. What is also important in this data is recognition of the lack of business support, lack of collaboration, and expertise.

Figure 45

What is Needed from Archive Systems to Assure Long-Term Readability?



This question tests if needs for long-term readability introduce any new requirements? Practitioners identified that they want migration solved, standardized logical formats, better repository systems, meta-data, integrity, & better management.

Several important points stand out in this graph:

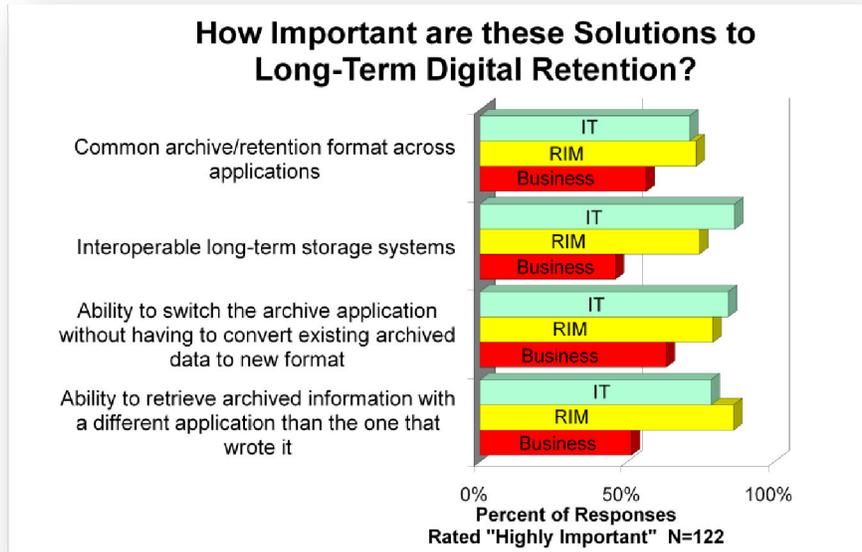
- 25% of respondents are looking for new technologies/solutions to the problem.
- 23% want higher data integrity, and better metadata and management tools (again, new technologies and solutions).
- 52% want better archive standards and hardware, migration tools, and information systems that address long-term retention needs.

100 YrATF Analysis: Unfortunately no one came forward with anything new. Rather, this is just a list of components to be included in a comprehensive solution. The conclusion still stands that the first technical goal is to solve the migration problems.

How Important Are These Solutions to Long-Term Digital Retention?

This question tests respondent’s expectations on how they would like to see the logical readability problem solved from three different viewpoints, IT, RIM, and the Business Group. The differences in responses illustrate the differences in job responsibilities well. IT wants better systems, RIM wants better long-term retention, and the Business Group just wants it taken care of.

Figure 46

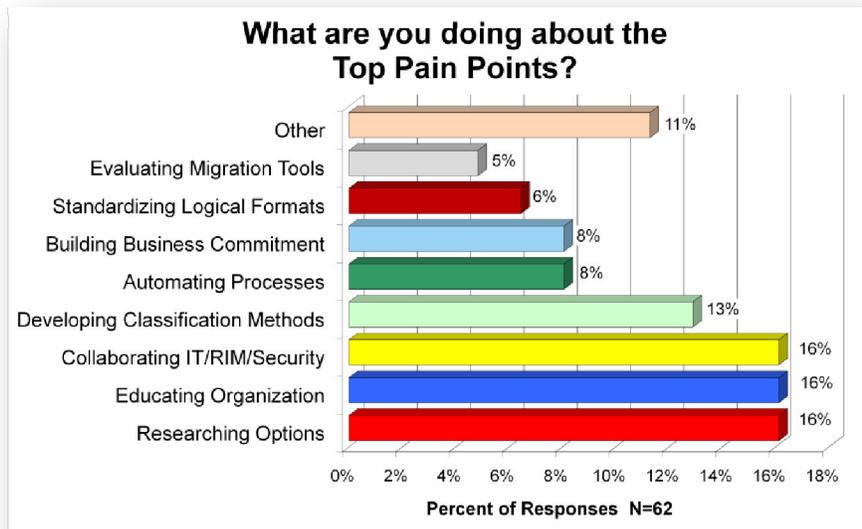


100 YrATF Analysis: These variances by job responsibility are normal and expected. Statistically, no one method of solving the problem seems to matter over the other.

What are You Doing To Address Top Pain Points?

During the discussion of problems and solutions, the survey not only asked what the “top pain points” are but, what the respondent’s organizations are doing about them. The responses clearly show that organizations recognize the need to address retention issues.

Figure 47



The majority, 57%, of the responses lie in the area of improving operating practices such as collaborating, classifying, automating processes, and standardizing logical formats.

100 YrATF Analysis: Today, practitioner’s have no silver bullets in the form of technology solutions. They have to rely on best practices beginning with collaboration and information classification.

What are Examples of Causes of Recovery Failures?

Respondents made a long list of experiences they have had in which they failed to recover information from the archives. The

Figure 48

Typical Recovery Failures

- Failure mode: costly, painful, or too slow
 - Email – slow and expensive
 - Had to reprogram the application to recover
 - Inability to timely answer audits – resulting in high cost, penalties, or unqualified audits
 - Bad CDs or tape
 - Hunting through media, incompatibility with search tools
- Quantified Losses
 - \$250,000
 - Too many to list
- Failure mode: inability to recover or read the information
 - Legacy information with decommissioned applications and/or systems
 - Occasional tape errors
 - Migration missed records or mappings to records
 - Old formats no longer supported
 - Unable to find records or locate tapes due to misfiling or erroneous search results
 - inability to read, corruption of data, & crashing of indexes
 - Lost attachments

examples fell generally into two categories, when the ‘failure mode’ was excessive cost, time, or pain and when they were unable to read or recover the information.

100 YrATF Analysis:

This list is interesting to read and further validates the complexity of the current operating environment plus the real need to solve the problems that today’s practitioners live with.

RESPONDENT COMMENTS:

As of yet we've had no experiences with recovering information. However, concerns to be aware of include the number of erroneous hits (applicability of returns in regards to search criteria), unable to locate the document due to misfiling; error in record as the application is unable to read the record; corruption of data in the archives; crashing of the knowledge base; stability of network (getting it back up after it goes down).

We have lost very little in 40 years and mostly derivative formats rather than archival or original.

Failures not experienced but anticipated include moving data to tape, then not being able to read the tape years down the line due to the proprietary nature of backup systems, especially if backup systems have changed in the interim.

**Respondent
Comments**

Examples of Failures

Outlook e-mails that were archived to a KVS repository lost the attachments when filed into the RMS (Records Management System).

What Caused The Failures? What Would You Do Differently Next Time?

This is the follow-on question to the previous one. The responses are tabulated and the comments listed for reading.

RESPONDENT COMMENTS:

Figure 49

Cause: Index file corruption in regards to the knowledge base. Recommendation: Test index files for corruption (and maintain back-ups)

Cause: Obsolete media. Media breakdown (diskette or CD went bad). Recommendation: I would tell the departments that had this happen to use mass storage devices, not optical media for storage.

Cause: Human error sometimes, lack of failsafes in procedures. Recommendation: Audit processes regularly, automate as much as possible, focus on archival formats.

Cause: Email Attachments - The process of reconnecting the attachments with the referenced stub failed during the filing process. Recommendation: Have IT test a product fully with the RMS integration before implementing.

Cause: Information even a few years old can no longer be read on Disks. Lack of training for users of the media. Recommendation: Partner RIMs and IT, ensure this is NOT an adversarial relationship (both sides need to work on this). Train IT in RIM methods early, at the basic level for IT. Basic RIM training for all staff, and better training for help desk staff. Ensure everyone understand every application best practices.

Analysis of Recovery Failures

What caused the failure?

- Index lost or corrupted
- Obsolete media, corruption
- Human error, poor procedures
- Change of application format
- Obsolete application or format
- Vendor no longer in business
- Lost tapes
- Migration failures
- Manual indexing, poor records
- System doesn't support discovery

What would you do differently?

- Test, audit media, check indexes, annual assessment
- Use disk, not tape
- Use an integrated system
- Collaboratively (IT & RIM) set classifications and requirements
- Make copies on microfilm
- Train staff, control and instrument process
- Set standards for file formats and repository interfaces
- Migrate information sooner

Respondent Comments

What Caused Failures

Respondent
Comments

What Caused Failures

Cause: Loss of substantial information when shifting applications (not well mapped and migrated). Generally caused by software interpretive rules and technology changes. Recommendation: I don't know what else I could have done

Cause: Bad indexing/listings and labeling of tapes in off-line storage, lack of regular quality checks of data tapes, not making back-ups and storing them separately, lack of data extraction testing from data tapes when processes are implemented. Recommendation: Include data model used for data on tape on all tapes. So, reading/understanding data can be achieved, especially if data tables or databases are changed over time. And, have good inventorying process for all off-line media.

Cause: Incompatible formats. Recommendation: At this point, we're mostly dealing with records written in formats that are no longer supported and/or available to us.

Cause: System unsearchable. Recommendation: Do not allow IT to determine system requirements with regard to RIM needs.

Cause: Unknown systems, lack of complete set of tapes, tape failures, data corruptions. Recommendation: Develop better systems, develop migration strategies, contain costs associated with migration, and have a better interface with RIM.

Cause: Human error and failure to adequately identify information and need to migrate it. Recommendation: Annual/periodic assessment of systems and full inventory of systems and the data they support. Regular meetings with data owners to ensure systems are still required and supported.

Cause: Loss of support from vendor. Recommendation: Backup on microfilm.

Cause: Outdated version, that was no longer accessible with today's technology. Recommendation: Would like one electronic database for use across the company that is updated and maintained by RIM in conjunction with Legal and IT.

Cause: Electrical, restore procedures. Recommendation: Go vanilla, don't built your own mousetrap, don't depend on non-scalable solutions, don't under-resource the training requirements for end-users.

Cause: Probably poor media. Recommendation: Buy certified tapes. Exercise tapes before writing Test before and after writing. Sample signal during retention period.

Cause: Absence of email management software. Recommendation: We are implementing email management software in all our major business areas. This is the one area

where the business case for storage costs can be made based on de-duplication, and where executives are aware of court cases impact on costs and reputation.

Cause: Lack of standardization for data entry and data preservation (unable to search and data corruption). Recommendation: Standardize practices for indexing. 'Force' vocabulary control. Standardize system administration. Usable metadata documentation. Adherence to preservation practices. In general, compliance with the agency's RIM policies and practices.

Cause: There are always records that don't migrate, usually from acquisitions. Different applications and platforms than the ones we use. Recommendation: We get IT teams in to do the migration for us.

Cause: Inability to timely answer audit questions resulting in unqualified audits. No structure or organization to the retention program, making the data that could be retrieved undependable. Recommendation: Start with the basics of putting policies into place and then enforcing them from the top down.

Cause: Logical Format and storage systems. Recommendation: KISS. Work with people first, establish rules and then implement systems.

Cause: I.T.'s failure to preserve e-mail. Recommendation: Printout e-mail and classify.

Cause: Too many to list. Data not indexed. Recommendation: Apply classification to all information and store according to class.

Cause: Media degradation. Recommendation: Do more output to microfilm.

Cause: Information in an old format. Recommendation: Migrate information sooner.

Cause: Data cannot be found; data take a long time to find. Poor storage system and poor storage practices. Recommendation: Nearline storage integrated with the ERMS.

Cause: Not able to find records through search tool. Poor indexing which is done manually. Primarily human performance errors. Recommendation: Constant training/review of importance of indexing and records, firmer controls for consistency in a database.

Cause: Db's produced with antique applications. The wetware refused to upgrade when it would still have been possible to

Respondent
Comments

What Caused Failures

**Respondent
Comments**

What Caused Failures

migrate the databases. Recommendation: Mirror the applications and databases, upgrade centrally. Be persistent.

Cause: Change server system and be unable to access data in the old format, types and logical format. Recommendation: Migrate all the information before changing.

Cause: Obsolete tape formats, no readers available. Inability to locate backup tapes, outdated email application formats, digital preservation planning. Recommendation: Including digital preservation & e-records requirements in the planning/design phase of upgrading or new systems procurements/development. Include archivists and records managers on enterprise architecture committees. Adopting open architecture and standards for file formats and repository interfaces.

What Recommendations Do You Have for Your Peers?

The last survey question gave respondents an opportunity to make recommendations. Their comments are summarized in the figure and listed for reading. The messages amplify and reinforce the conclusions drawn and presented in the executive summary of this report.

Figure 50



RESPONDENT COMMENTS

When using a digital archive understand you will have a long hard expensive road to keep the records. You have to think about the ability of your great, great, great, great ... grandchildren being able to read and logically interpret what your history was.

Remember that IT doesn't own the information. RIM, Legal, Business units and IT all have a part to play in the decisions

applied to business records and should be sitting down at the table together.

Print to paper if possible and manage until a real solution comes along.

Metadata is important, and it's better to implement the metadata at the front end rather than populating the metadata after the record has been saved into the repository. Work in tandem with IT.

I would tell the departments that had this happen to use mass storage devices, not optical media for storage. Backup your data on a regular basis. Refresh and migrate records on a regular basis. Call me if they're having problems. Just a general comment on the survey - the vocabulary used in it seems problematic. Records retention is different than depositing something in an archives. Archiving is a very problematic word and I would suggest not using it. It suggests dumping records into some bottomless pit where they can be forgotten. Ingest into a recordkeeping environment, or to permanently preserve for long-term records retention seem better. Long-term seems problematic as well, because you haven't defined it - does this mean greater than 25 years or greater than 100 years, or the day after I retire?

Collaborate and rely on standards and good practice.

Talk to Archives. They have been looking at this problem longer than business has. Involve senior management early in the process and be sure there is a common goal outside the influence of technology enthusiasts.

Prayer. Backup on microfilm.

Get RIM and IT at the same table. Create a relationship. Both need to be included up front to develop solutions that will work on both sides.

Inventory all systems and data Determine who owns and uses what Pay attention to organization restructuring and how it impacts data Budget to support and convert data in systems when systems are initiated and/or deployed Greater involvement between IT and RIM regarding retention requirements and segregation of data based on privacy and other issues A clear understanding between IT, Users and RIM of terms 'Archiving', 'Retention', and 'Backup' to ensure they are properly applied to management of information.

For absolute produce-ability maintain a hard copy or create a secure/backed-up/updated electronic repository using pdf images.

Stay with industry standards. THINK about why you're retaining something. Just THINK about it. And don't spend a dime to save a nickel. Keep as little as possible in long-term.

Respondent
Comments

Peer
Recommendations

**Respondent
Comments**

Peer
Recommendations

Research your legal requirements and use technology as fully as possible to classify retention periods and destruction.

Communication and visibility about the issues Compliance with agency's RIM policies and practices – enforcement.

Upper management advocacy.

If you are fortunate enough to be in the early years of your data, get on top of it now. It is much easier to keep the data tame from the beginning than to try to tame it later.

Balance claims of vendors & IT for long-term accessibility with recommendations of records managers.

KISS - work with people first, establish rules and then implement systems.

Preserve hardcopy in offsite storage wherever possible and feasible.

Research, read and participate in training, seminars, and ARMA/AIIM events.

Segregate your information by record type.

Output to Microfilm. Index well and output to the only true 100 year media.

Know that this has to be addressed frequently. You cannot assume someone else is taking care of this.

The customer does not get to determine the policies of the RM department.

Write a plan. Ours is done this year, 2006, approved as of Jan 2007. Next, get it implemented. All archives to be in order from 2007 to 2012 when all must be in place.

Keep as little as possible in long-term.

-- End --

APPENDIX

About the Authors

Michael Peterson: is Chief Strategy Advocate for the SNIA's Data Management Forum and President of Strategic Research Corporation where he leads a market and business development consultancy and information technology research practice. Michael is the founder of the SNIA and was the past president from 1998 to 1999 and a co-founder of the Data Management Forum in 2003. He is a well-know international speaker and visionary.

Gary Zasman: is the Co-Chair of the 100 Year Archive Task Force and is the WW Practice Director of Network Appliance's assessments and optimization services. In 2001, a team that Mr. Zasman worked with (over a six-year period) was selected as a finalist for the prestigious Computerworld Smithsonian Award, for developing a digital archive and associated processes to house and manage hundreds of terabytes of unstructured data and associated metadata for historical preservation.

Peter Mojica: is the Co-Chair of the 100 Year Archive Task Force and is Vice President, Product Strategy and Management of AXS-One. A 20+ year Information Technology veteran of the financial services and software industries, Peter leads AXS-One's compliance solutions strategy for digital archival, long-term records management of electronic records, regulatory compliance and legal discovery with a focus on electronic communications such as e-Mail, instant messaging, office documents and transactional data for worldwide enterprises.

Jeff Porter: is the Chairman of the SNIA's Data Management Forum and a Senior Technologist in the Office of the CTO of the Information Management Software Group at EMC where he is an advocate for information-centric management practices. He has over twenty-five years experience in information technology disciplines and is responsible for innovating new product solutions using existing, evolving, and emerging technologies at EMC.

The Authors

Michael Peterson
Jeff Porter
Gary Zasman
Peter Mojica



**100 Year Archive
Task Force**

[www.snia-
dmf.org/100year](http://www.snia-dmf.org/100year)

About the 100 Year Archive Task Force

The SNIA's 100 Year Archive Task Force is a global, multi-agency group working to define best practices and necessary storage standards for long term digital information retention and preservation. All interested parties and organizations are invited to join in this work effort. The Task Force is a project sponsored by the SNIA's Data Management Forum. www.snia-dmf.org/100year

About the SNIA

The Storage Networking Industry Association (SNIA) is a not-for-profit global organization, made up of more than 460 member companies and close to 7,000 active individuals spanning virtually the entire storage industry. SNIA members share the common goal of advancing the adoption of storage networks as complete and trusted solutions. To this end, the SNIA is uniquely committed to delivering standards, education and services that will propel open storage networking solutions into the broader market. For additional information, visit the SNIA web site at <http://www.snia.org>.

About the SNIA's Data Management Forum

The SNIA Data Management Forum is a cooperative initiative of IT professionals, vendors, integrators, and service providers working together to conduct market education, develop best practices and promote standardization activities that help organizations become Information-Centric Enterprises. Areas of focus include the technologies and services that support information lifecycle management, data protection, information security, and long-term digital information retention and preservation. For more information, visit www.snia-dmf.org.

