



Education

# **Best Practices for Long-Term Retention & Preservation**

Michael Peterson, Strategic Research Corp.  
Gary Zasman, Network Appliance

# SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - ◆ Any slide or slides used must be reproduced without modification
  - ◆ The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.

## ➤ Best Practices for Long-Term Retention & Preservation of Digital Information

- ◆ Compliance, legal, business, discovery, and security risk have changed the value and risk of owning and administering information within the datacenter. Old approaches to retaining, preserving, and disposing of information in multiple isolated and 'siloes' 'electronic archives' no longer meet today's requirements for reduced operating costs, scale, and high efficiencies. This presentation highlights new work being spearheaded by the SNIA's Data Management Forum to produce a reference architecture for best practices in long-term digital information retention based on information-lifecycle-management methods.

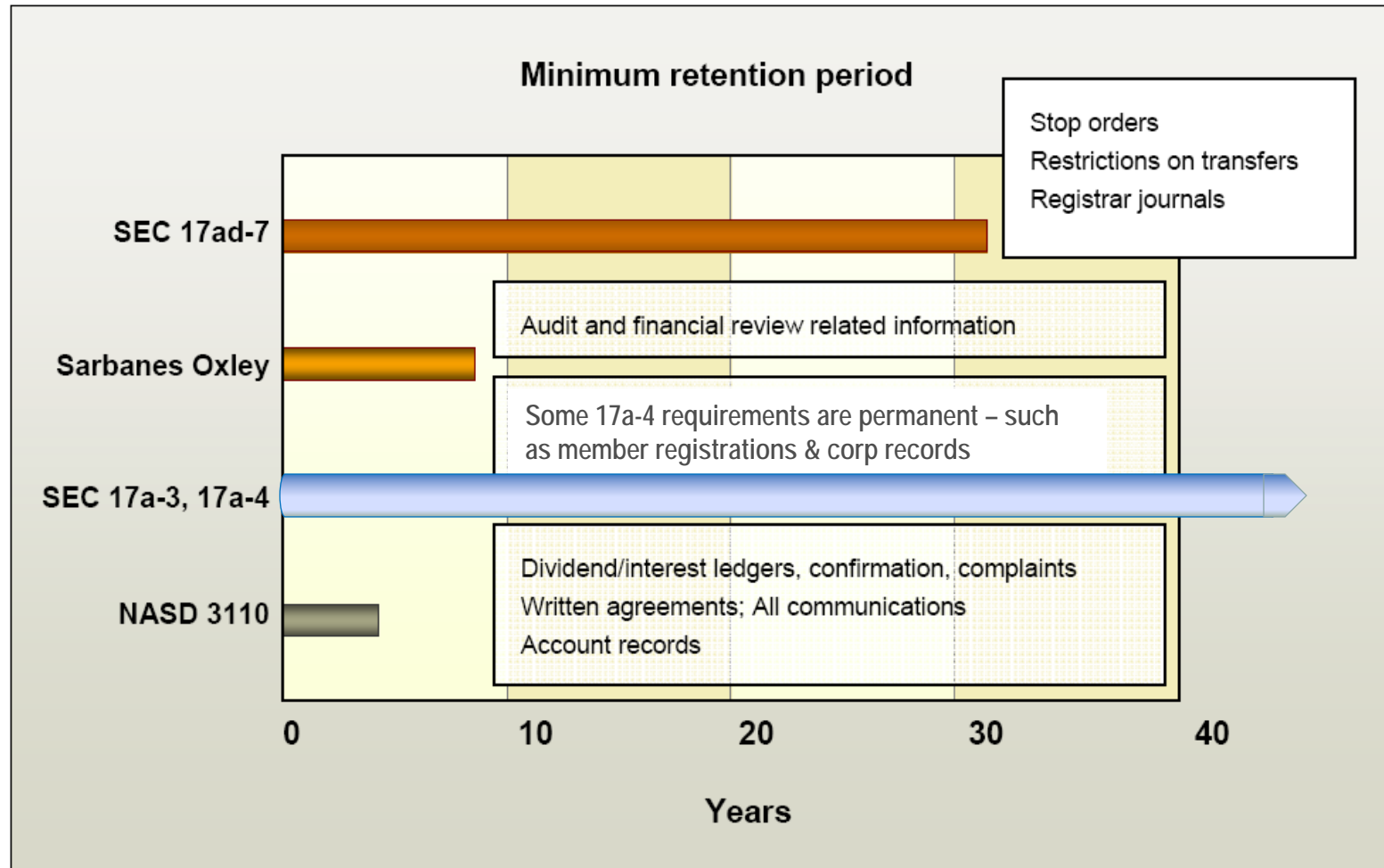
# Why do You Care?

- Real Drivers: Legal, Security, Compliance, and Business Risk
  - ◆ eDiscovery, Compliance, Legal Hold, Customer Privacy, Business Loss, Fines, Theft, Damage...
- Impact: **More information is being kept long-term than you think – and it is at risk!**



Long Term Parking  
1982 - Arman

# Long-Term is Real!



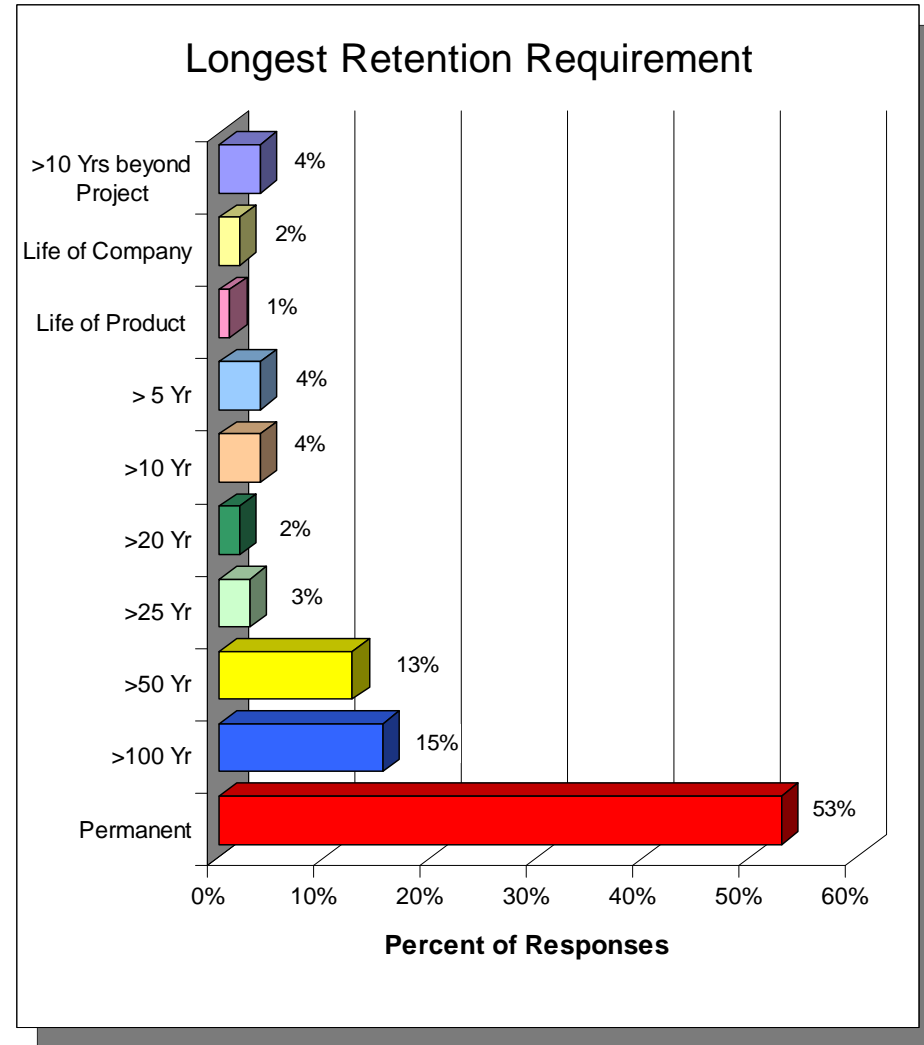
Retention Requirements for Financial Services Companies

# Long-Term is Real!

## ➤ Longest Requirement

- ◆ 68 % over 100 Years
- ◆ 83% over 50 Years


## ➤ Requirements vary by organization type, information type, and compliance rules/risk



Source: 100 Yr Archive Requirements Survey 2007 , N=104

# Long-Term is Real!

<b>STATE OF TEXAS</b>		<b>SLR 105</b>
<b>Records Retention Schedule</b>		<i>Form SLR 105C must accompany this form.</i>
Approved <u>9/26/06</u>		1. Page 82 of 273

2. Agency Code <b>745</b>		3. Agency Name <b>The University of Texas Health Science Center at San Antonio</b>						
4. Record Series Item No.	5. Agency Item No.	6. Record Series Title	7. Retention Period			8. Archival	10. 106 No.	
			Agency	Storage	Total	9. Remarks		

3.1.014	180 EE, DE, PR	Employment Selection Records - Faculty and Executive Committee Members. Includes notes of interviews with candidates; questions asked of applicants; audio and videotapes of job interviews; driving record and previous injury checks; pre-employment physical examinations; polygraph examination results; sanction checks; waiver requests; special circumstance requests; selection of candidate forms; cover sheets for faculty, fellow and staff appointments; and all other records that document the selection process.	AC+5, 75	<span style="border: 2px solid red; border-radius: 50%; padding: 2px;">AC+5, 75</span>	AC = Recruitment date (as per OFCCP - Office of Federal Contract Compliance Program - requirements). EE keeps faculty recruitment reports 75 years. Sanction (Office of Inspector General/General Services Administration (OIG/GSA) Database Check), security, and driving clearances are electronic. Previous injury checks and physical exams are held by departments. All records for Fellows are held by departments with notification to EE. President's office may retain records for applicants for Executive Committee members and department chairs for as long as they are administratively valuable.
---------	----------------	---	----------	--	---

Retention Codes (Field 7)	Archival Codes (Field 8)
AC - After Closed, Terminated, Completed, Expired, Settled AV - Administrative Value	I - Retain in University Archives O - Review by University Archives
CE - Calendar Year End FE - Fiscal Year End	LA - Life of Asset MO - Months
PM - Permanent US - Until Superseded	

# Impact of Long-Term Retention

- Real Drivers: Legal, Security, Compliance, and Business Risk
  - ◆ eDiscovery, Compliance, Legal Hold, Customer Privacy, Business Loss, Fines, Theft, Damage...
- Impact: More information is being kept long-term, and it must be accessible and discoverable
  - ◆ Lack of confidence that information can be retained long-term



# Key Survey Data Points

- 70% of respondents say they are ‘highly dissatisfied’ with their ability to read their retained information in 50 years
- Current practices are too manual, too prone to error and too costly
- Collaboration is recognized as necessary in order to define information retention requirements



Source: 100 Yr Archive Requirements Survey 2007 N=276

*“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)*

# Impact of Long-Term Retention

- Real Drivers: Legal, Security, Compliance, and Business Risk
  - ◆ eDiscovery, Compliance, Legal Hold, Customer Privacy, Business Loss, Fines, Theft, Damage...
- Impact: More information is being kept long-term, and it must be accessible and discoverable
  - ◆ Ever increasing cost and complexity
  - ◆ Requirements for permanent deletion, discovery, authenticity, and other preservation services
  - ◆ How are you going to migrate petabytes of information per year?

# Migration is a Requirement

## ➤ Physical Migration

- ◆ Moving information from one physical system or location to another or from one physical media-format to another to maintain physical readability, accessibility, and integrity or to achieve other storage benefits
- ◆ Drivers: media failure, media or system obsolescence, system changes, cost of operations (people, power, space), long retention requirements
- ◆ Inhibitors: cost, complexity, sheer volume of information, lack of budget, lack of time...

# Migration is a Requirement

## ➤ Logical Migration

- ◆ Moving information from one logical-format to another, such as from an old application version to a new version, to preserve readability, interpretability, and integrity.
- ◆ Drivers: changing application formats, obsolete applications, mergers, increasing retention requirements
- ◆ Inhibitors: cost, complexity, time, expertise, volume of information, non-standard formats, poor practices, lack of budget, time, people...

- Logical and physical migration do not scale cost-effectively
  - ◆ Only operating standard today is to migrate information physically (to new media) every 3-5 years and logically (to new formats) before the applications and readers die and become obsolete (every 5-10 years)
    - › **A never ending, costly cycle of migration**



# Survey Conclusion

- Logical and physical migration do not scale cost-effectively
  - ◆ Practitioners are struggling to keep up with migration requirements. Only 30% claimed to be doing physical migration correctly on disk & none on tape or optical. Only 20% claimed they were confident in their ability to logically migrate some of the data.
    - › **Information is at risk long-term!**



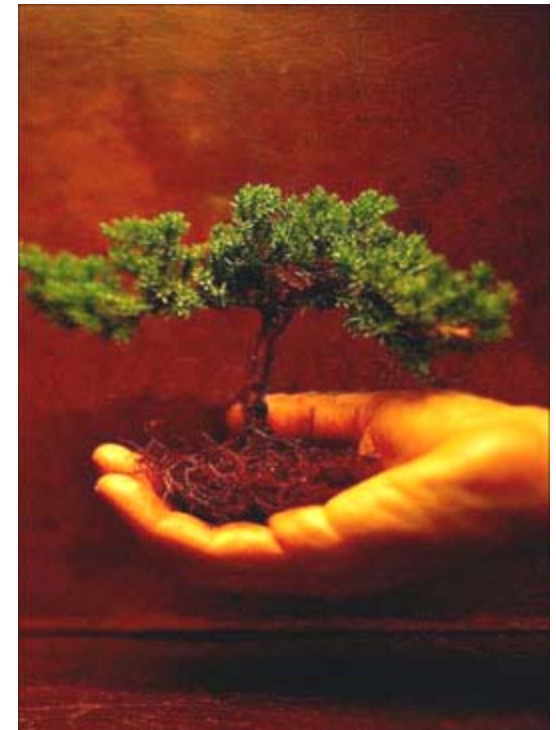
# Best Practices for Long-Term Retention & Preservation



# We Need a Holistic Approach

Strategy: Apply IT Service Management & Information Lifecycle Management

- Move away from ‘unmanaged repositories’ and silos of ‘disconnected information’
  - ◆ Use preservation-aware applications
    - › ILM-based practices and ILM-capable repositories
    - › Focus on “retention and preservation” not creating silos



# Retention and Preservation

## ➤ Retention:

- ◆ To keep and control information objects for specific periods of time. As a policy, a retention period for an information object or class of objects specifies the period of time the information is to be retained along with its administrative, legal, fiscal, historical, business, security, or other disposition requirements. (Source: SNIA and Society of American Archivists)

## ➤ Preservation:

- ◆ The processes and operations involved in ensuring the technical and intellectual survival of authentic information objects through time. (Source: NARA-ERA)

- A property of an information object's content and metadata that identifies that it is currently what it was originally and verifies that its content has not changed over time. (Source: SNIA, National Archives & Records Administration, Society of American Archivists)
  - ◆ To maintain authenticity requires maintenance of the information object's digital integrity through preventing change or corruption, verification that it is the original, auditing access, and providing a means to detect change typically accomplished through reliable hashing methods, security, and audit trails.

# Why is Preservation a Problem?

- Who cares?
  - ◆ Archive responsibility is at the bottom of the IT hierarchy and lacks adequate funding
  - ◆ Not associated with a business opportunity – rather in mitigating risk (insurance)
- Drivers are relatively new (compliance, legal...)
- Technologies are incomplete and immature
  - ◆ Archivists rely on intensive care and best practices – these approaches don't scale to the datacenter
- Failure to Collaborate (isolated responsibilities)

# IT Preservation Practices

- What are the requirements? (Most do not know!)
- Many still rely on Backup (Wrong!)
- Record to tape and 'lose it' (Sad but true!)
- Migration by Crisis:
  - ◆ Only 30% Migrate every 3-5 years if on disk, 0% migrate regularly if on tape, if an application changes, it forces a 'crisis' migration



# Records Management Retention Methods

## ➤ Developing an Electronic Records Retention Schedule

- ◆ Conduct an electronic records inventory
- ◆ Conduct legal research to obtain regulatory and legal retention requirements
- ◆ Work with various organization members, business, legal, compliance, & security retention requirements
- ◆ Identify vital records & publish, educate, and implement

ARMA International				
RECORDS RETENTION AND DISPOSITION SCHEDULE				
Listing by Department				
Records Series Code	Records Series Title	Responsible Department	Total Retention Period	Vital Record?
02.010000	Activity Reports Committees Headquarters	CORPORATE	3 years	
04.010000	Administrative Letters	CORPORATE	10 years after superseded	
04.020000	Articles of Incorporation	CORPORATE	Life of Association	Yes
04.030000	Bylaws	CORPORATE	Life of Association	Yes
02.062000	Certificates of Destruction	CORPORATE	10 years after records are destroyed	
04.040000	Charters Chapters Regions	CORPORATE	Life of Association	Yes
04.050000	Contracts/Leases	CORPORATE	6 years after contract expires	Yes
04.060000	Copyrights	CORPORATE	Life of Association	Yes
02.030000	Correspondence (General) Ad Hoc Associations (Other) Chapters Committees & Publications Headquarters Officers Regions	CORPORATE	3 years	
02.031000	Elections/Nominations Ballots List of Elected Officers	CORPORATE	3 years	
02.035000	Email Messages Inbox & Sent Items Insurance Policies Directors/Officers Liability Property	CORPORATE	90 days maximum 6 years after contract expires	Yes

# Records Management Retention Methods

## ➤ Electronic Records Preservation Processes and Controls

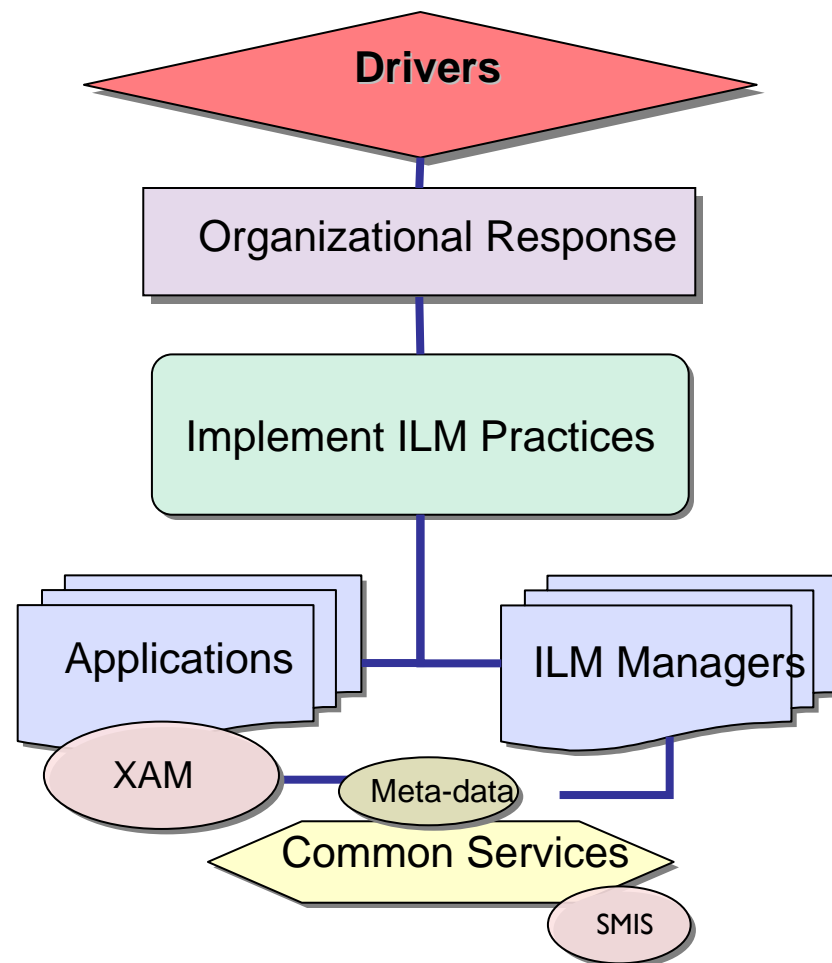
- ◆ Appraisal
- ◆ Ingest
- ◆ Storage
- ◆ Preservation actions
- ◆ Access
- ◆ Disposition



**It Takes More Than Setting Policies**

## ➤ Collaborate

- ◆ Stakeholders: IT, RIM, Legal Business, Security



- “Classification: The Cornerstone for Information Management”
- “The Secret Sauce of ILM”

\* ILM = Information Lifecycle Management

# The Collaborative Team

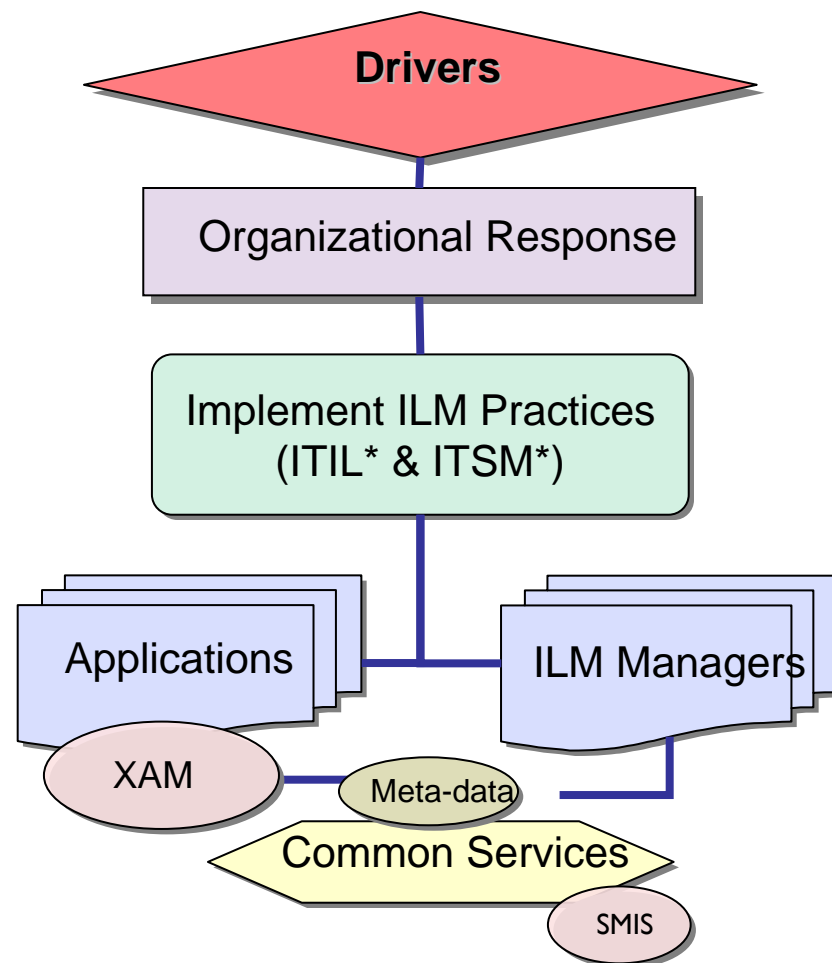
- **Legal** – Assist in the identification of information and its importance to the organization from legal, business, and compliance perspectives plus the definition of security, retention, and compliance policies
- **Records Managers** – Evaluate policy and procedures periodically, analyze current risks, implement regular reviews, determine retention requirements
- **IT** – Implement the policies and define systems for storage and security of digital information and records including metadata, log files, and audit trails

# The Collaborative Team

- **Business/Operations** – Responsible for the creation, receipt and storage of active records and metadata as part of their job duties, in accordance with established policies and procedures including definition of business requirements
- **Security** – Responsible for assisting in defining security, confidentiality, and compliance policies and in their implementation and auditing
- **Archivists** – Responsible for preservation of digital and analog assets – use them and tap their experience

- Collaborate
  - ◆ Stakeholders: IT, RIM, Legal Business, Security
- Identify Assets & Resources
- Classify Information
- Set Requirements
- Implement Services
- Measure and Improve

- ILM = Information Lifecycle Management
- ITIL = Information Technology Infrastructure Library
- ITSM = IT Service Management



# Example Helpful Practices

- Classify your information (into a few common buckets)
- Set retention periods and delete 'expired' information
  - ◆ Free up space, only store what is required
  - ◆ Include your databases
- Control the number of copies for protection and operational recovery and their locations
- Set policies for audits and perform them
  - ◆ Measure and improve



Storage Considerations for  
Database Archiving

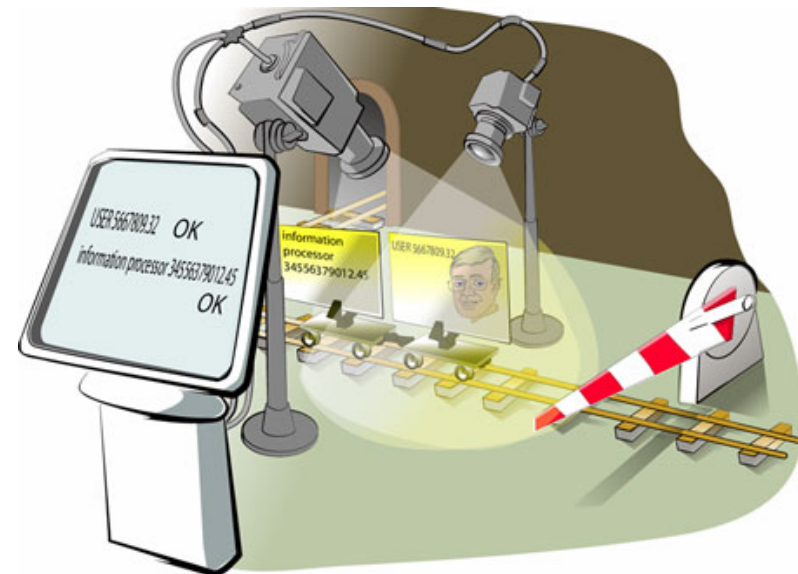
# Preservation Policies

- Long-term preservation policy should:
  - ◆ Identify the business, legal, and compliance goals.
    - › For example: retention periods of classes of information, requirements for integrity, authenticity, auditing, security, confidentiality, accessibility, ...
- A description of best practices to which the storage repository adheres including migration

***Collaborate, Identify, Classify, Requirements***

# Preservation Process Controls

- Define Preservation Metadata Requirements
  - ◆ Rules for capturing & preserving metadata should be incorporated into all business process procedures
  - ◆ Sample policy elements:
    - › Determine the length of time that records will be preserved
    - › Organization's risks requirements
    - › Logical Migration plan

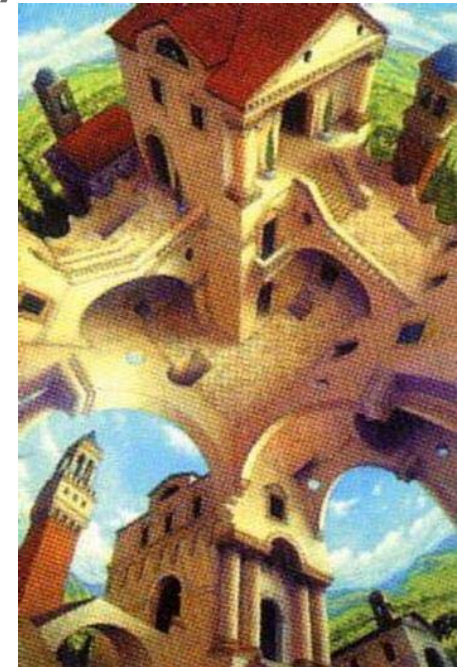


# Preservation Process Controls

- ◆ Storage Media – The media must ensure readability, integrity, and authenticity for as long as needed.
  - ◆ Media must be protected from unauthorized access, loss, tampering, destruction, theft, disaster and be discoverable.
  - ◆ Must be physically and logically migrateable
  - ◆ Media types with the right attributes:
    - › Disk – Content Aware Storage, Write-Once-Read-Many, or via digital signatures, hashing, etc.
    - › Tape – WORM
    - › Optical - WORM

## RECOMMENDATIONS

- Move from fixed-term (3-5 yrs) to ‘when needed’
  - ◆ 1. Today: Self-healing systems (disk, tape, optical)
  - ◆ 2. Goal: ‘federate and virtualize” these systems
- The only physical ‘migration’ that remains is when changing or “moving” platforms, ‘upgrading’, ‘replacing’...



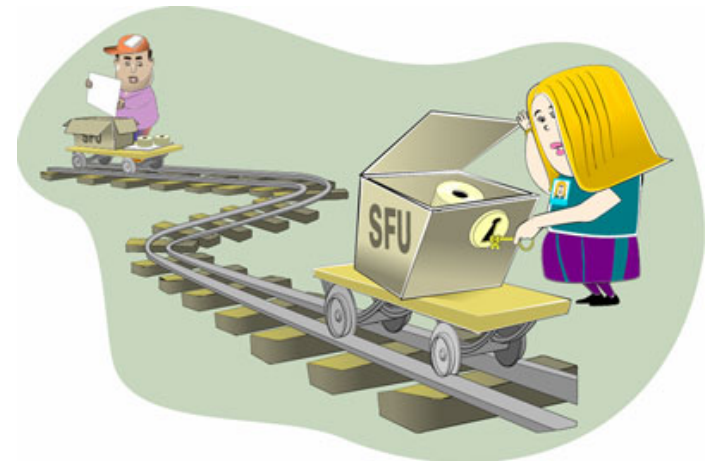
- Move to a repository system that provides preservation features:
  - ◆ Unique naming, de-duplication, integrity, discovery, authenticity, security, confidentiality, business continuity, etc....
  - ◆ However – no interop standards for de-dup, WORM, or Content Aware systems
    - › Until XAM\* is finalized and adopted



\* *XAM = eXtensible Access Method, an application to storage interface standard in development within SNIA*

# Logical Migration

- The ability to read and interpret information must be updated as applications and readers obsolesce
  - ◆ Must be able to maintain authenticity through a migration
  - ◆ In some cases, only the appearance needs to be maintained, in others the entire content
    - › A spreadsheet with macros, links, and formulas has hidden information of value



# Logical Migration

## ➤ Methods

- ◆ Use the application to ‘update’ old data or files
- ◆ Encapsulation – Retains the records in its original form, but encapsulates it with a set of instructions on how the original should be interpreted. (XML wrapper)
- ◆ Emulation – Using a device or program in place of a different one to achieve the same effect as the original.
- ◆ Transformation – into a standard format (example, an image of a document – tiff or pdf-a)

# Logical Migration - Today

## ➤ “Best Practices “

- ◆ Set up a policy and procedure to protect the organization from upgrading, changing, and obsolescing applications and readers
- ◆ Migrate when needed
- ◆ Don't forget 'analog' retention if needed



## ➤ Transform

- ◆ Into a standard format (example, an image of a document – tiff or pdf-a)
- ◆ As standard as possible - (supported by multiple software applications and operating systems)

# Preservation Process Controls

- Security – Storage repositories and information are to be protected against inappropriate access, viewing, change, copying, or misuse
  - ◆ Auditable controls on access, privacy, integrity, and authenticity to ensure that information is not compromised.



# Preservation Process Controls

- Disposition – When the records retention requirement is met, all copies of records including series, preservation and backup copies should be permanently deleted
  - ◆ Define disposition methods and policies by class of information in advance
  - ◆ Permanent Deletion is a key method to reduce the storage load, the legal risk, and the cost

## ➤ People

- ◆ Train staff on key retention, preservation, classification, and security practices and technologies

## ➤ Process

- ◆ Use ILM-based practices:
  - Collaborate, identify, classify, requirements, implement, measure, and improve
- ◆ Have periodic audits on long-term retention stores
- ◆ Document logical & physical migration processes

## ➤ Technology

- ◆ Plan for and incorporate robust preservation services:
  - Meta-data, eDiscovery, deletion, security, authenticity, etc.
- ◆ Deploy storage systems that scale and self-heal

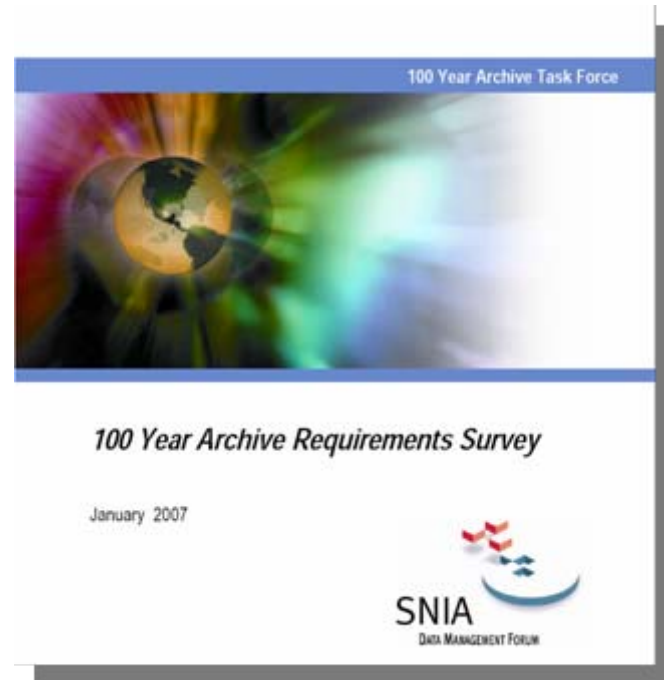
## ➤ Standards

- ◆ Transform or encapsulate into XML formats wherever possible
- ◆ Support SD-SCDF and XAM standardization

# SNIA-DMF Projects

## ➤ Long-Term Retention Reference Model

- ◆ Requirements (done)
- ◆ Glossary (done)
- ◆ Best practices for storage
- ◆ Define a reference architecture covering migration, security, authenticity, etc.



1. OAIS: Open Archival Information System

# Long-Term Retention Projects

## ➤ Logical/ Physical Migration and Movement

- ◆ Launch a TWG to define “SD-SCDF”, a self-describing, self-contained data format standard
- ◆ SD-SCDF provides a standard container allowing all key preservation attributes to be maintained over time and across virtualized repositories

## ➤ Conduct Market Education

- ◆ Speaking, papers, web
- ◆ Interact with international community working on retention and archive



## ➤ SNIA Data Management Forum

[www.snia.org/forums/dmf](http://www.snia.org/forums/dmf)

- ◆ 100 Yr Archive Task Force Requirements Survey
- ◆ “Terminology Bridge” White Paper & Glossary

## ➤ DMF Community

- ◆ <http://community.snia-dmf.org>
- ◆ A networking and collaborative community, working to create “information-centric enterprises”



- Please send any questions or comments on this presentation to SNIA: [trackdatamgmt@snia.org](mailto:trackdatamgmt@snia.org)

**Many thanks to the following individuals  
for their contributions to this tutorial.**

**- SNIA Education Committee**

**Michael Peterson, Chief Strategy Advocate, DMF**  
**Peter Mojica, Co-Chair LTACSI**  
**Gary Zasman, Co-Chair LTACSI**  
**Craig Mullins, Co-Chair LTACSI**