



Education

Solving the Coming Archive Crisis

Michael Peterson,
President, Strategic Research Corp.
Chief Strategy Advocate ,
SNIA Data Management Forum

- 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.

About SNIA and the DMF

About the Storage Networking Industry Association (SNIA)

- ◆ SNIA's primary goal is to ensure that storage networks become complete and trusted solutions across the IT community
- ◆ For additional information about SNIA see www.snia.org
- ◆ SNIA's "Dictionary of Storage Networking Terminology" is online at www.snia.org/dictionary

About the SNIA Data Management Forum (DMF) www.snia-dmf.org

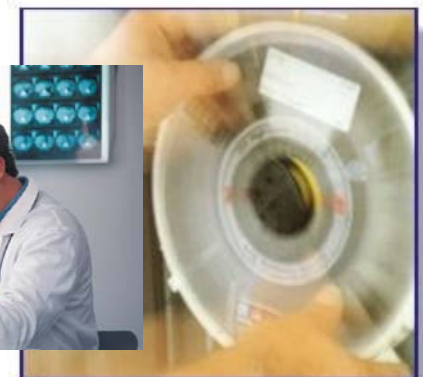
- ◆ The DMF is a sub-group of SNIA acting as the worldwide authority on Data Management, Data Protection, Preservation, and ILM
- ◆ The DMF is a collaborative storage industry resource available to anyone responsible for the accessibility and integrity of their organization's information.

DMF		
Data Protection Initiative (DPI)	Information Lifecycle Management Initiative (ILMI)	Long term Archive and Compliance Storage Initiative (LTACSI)
Defining new approaches and best practices for data protection and recovery	Developing, teaching and promoting ILM practices, implementation methods, and benefits	Addressing challenges in developing, securing, and retaining long-term digital archives

What is Wrong Here?

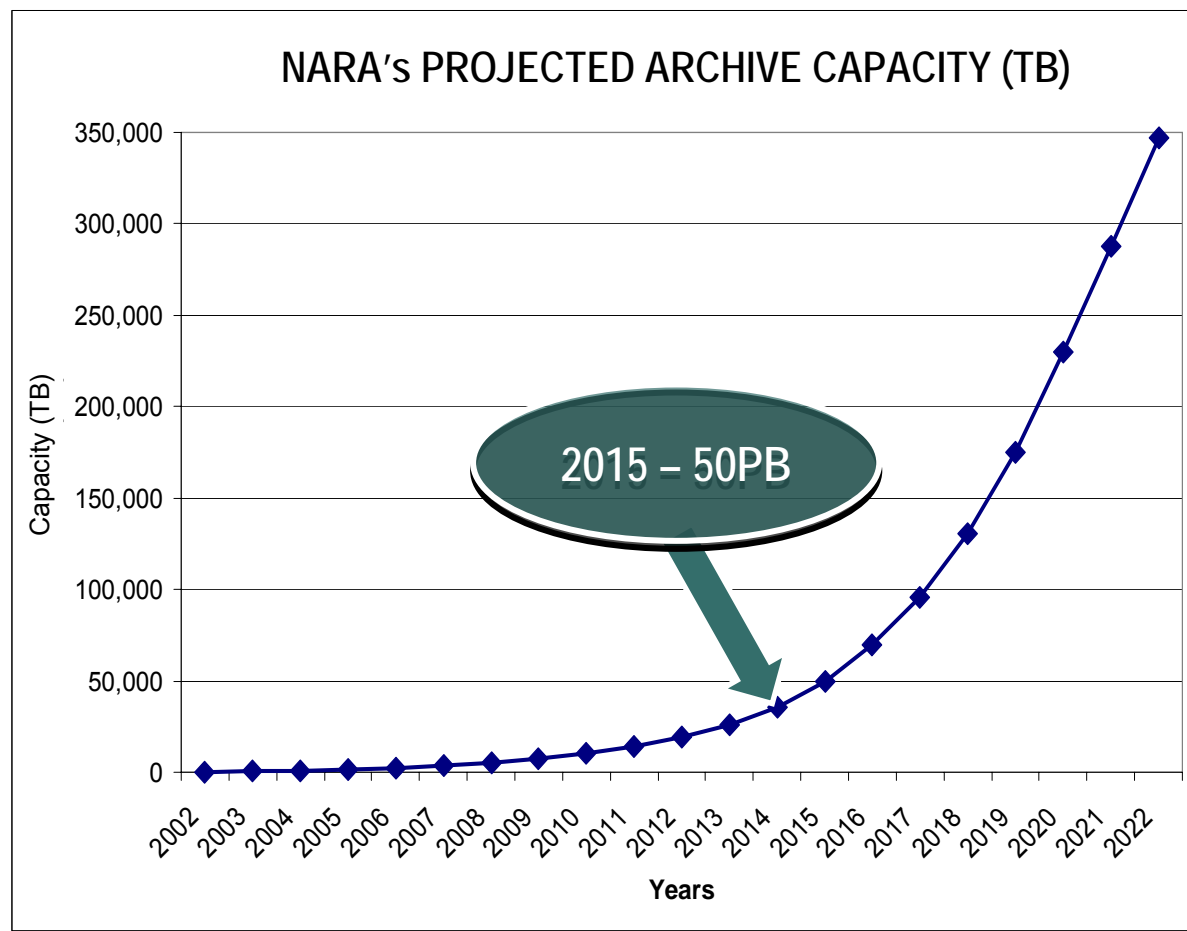
- *As an archivist, I oversee a collection that we've maintained for 40 years which includes published materials from as early as 1460 and manuscript materials from as early as 1390.*
- *In terms of data and automation system, the collections include Hollerith cards, Dictaphone tapes, 5 ¼ disks, 3.5 disks, CDs, DVDs, etc. We also maintain a small collection of 386 and 486 machines, drives, as well as various software packages (Windows 3, Wordperfect 4.2, 5.0, 5.1 , 6, etc.) in order to read and migrate and convert.*

Source: 100 Yr Archive Requirements Study, An archivist at a major University



Or Here?

- How do you keep up with the growth?
- How do you migrate >10 PB per year?
- How do you fund “never-ending” cost?
- How do you not lose information?



Source: Ken Thibodeau, Director NARA ERA Project

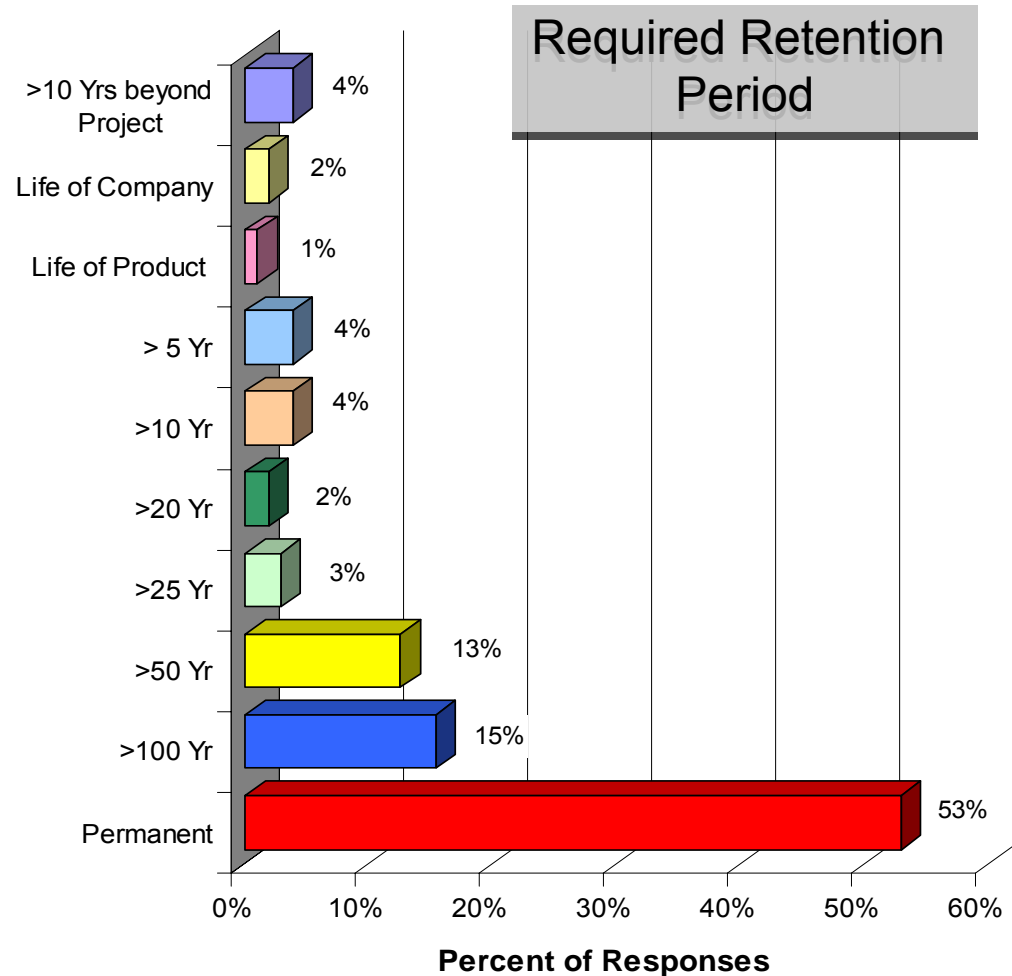
Or Here?

➤ For how many years is data migration needed?

- ◆ 81% of respondents responded over 50 years
- ◆ How will they do that?

➤ What is long term?

- ◆ A period long enough to requires logical & physical data migration
- ◆ > 10-to-15 years



Source: 100 Yr Archive Requirements Survey, 01/07, n=276

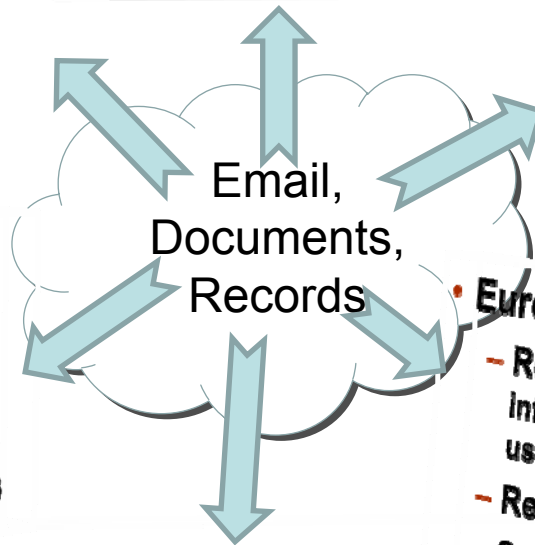
Retention Requirements are Real ^{Education} SNIA

- **7 year Retention**
 - Payroll
 - Inventories
 - Expense Reports
 - Safety Records
 - Payables
 - Receivables

- **15 year Retention**
 - IRS Audits
 - Tax Work Papers
 - Retirement History
 - Audited Finance Stmt

- **40+ year Retention**
 - Medical Records
 - Accident Reports
 - Pension Contributions

- **Permanent Retention**
 - Contracts & Leases
 - Legal Correspondence
 - Insurance Reports
 - Patent/Copyright Docs
 - Property Appraisals
 - Training Manuals
 - Union Agreements



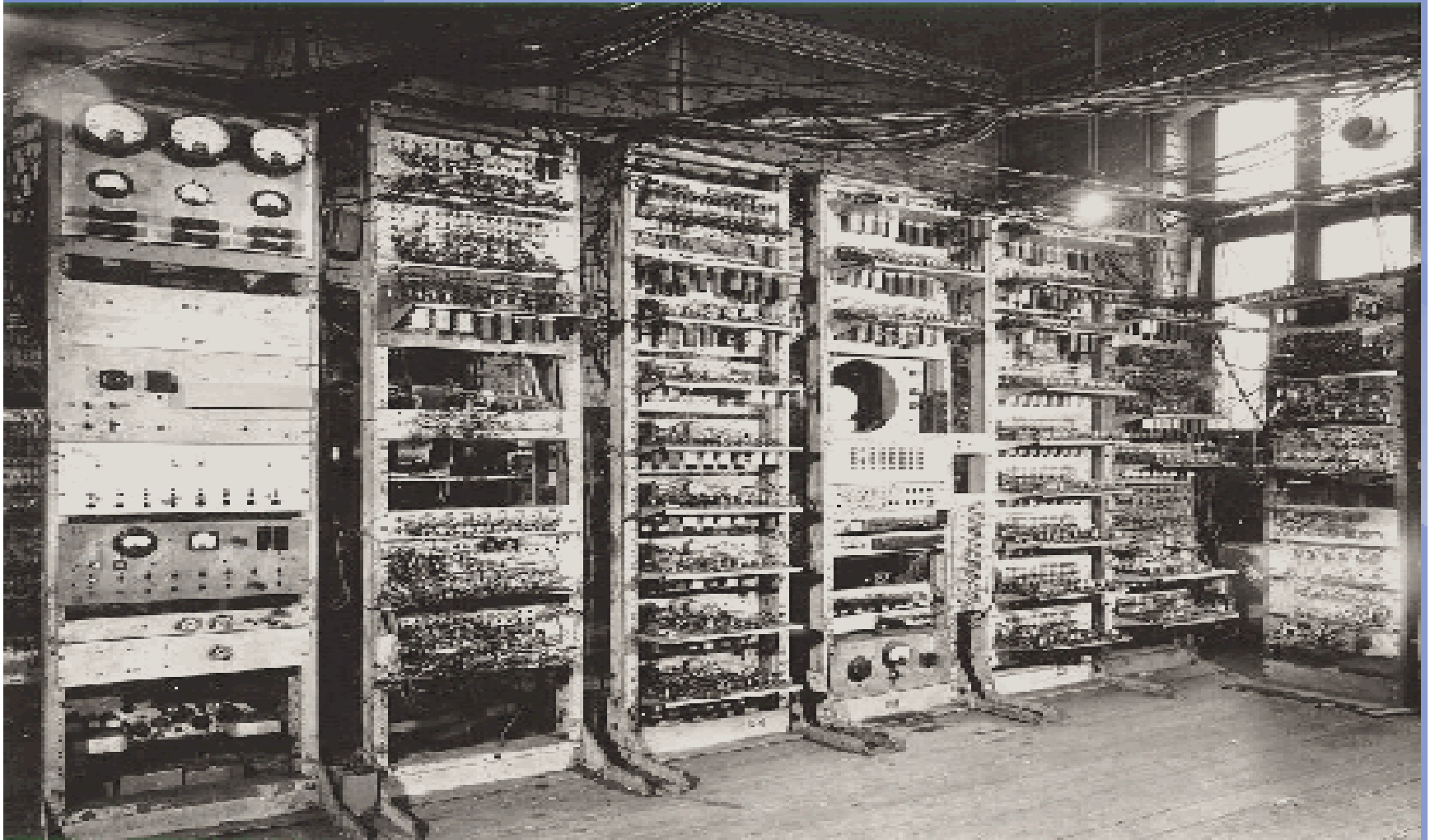
- **European Union – Dir. 2006/24/EC**
 - Retain **all electronic communications** info (Land, Mobil, VoIP) including phone #, user ids, equipment, etc.
 - Retain 6 month to **2 years**
 - Starting 3/2006 – Delayed until 3/2009

U.S. Army

- Support 2.4 Million Soldiers
- Retain **Life of Service + 62 years**

Does Your Business Know the PAIN of Preserving Information?

Education
SNIA



October, 2007

SNIA-DMF: Solving the Coming Archive Crisis
2007 Storage Networking Industry Association. All Rights Reserved.

➤ Technology Challenges

- ◆ Maintaining physical readability
- ◆ Maintaining logical readability
- ◆ Migration of large repositories
- ◆ Emulation of formats
- ◆ Maintenance of historical readers and applications
- ◆ Protection from change
- ◆ Protection from loss or damage
- ◆ Physical & logical security
- ◆ Automation
- ◆ Deletion & disposition
- ◆ Search – discovery
- ◆ Testing/auditing

➤ Operations Challenges

- ◆ Collaboration on and setting requirements
- ◆ Classification
- ◆ Establishing adequate Metadata
- ◆ Standardizing practices
- ◆ Finding value in the archives
- ◆ Professional recognition



Your Peers Agree?

*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.*

-- Archivist

*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.*

-- RIM professional

*The rate of change in computing technologies is such that **information can be rendered inaccessible within a decade**. **Preservation is therefore a more immediate issue for digital** than for traditional resources. **Digital information will not survive or remain accessible by accident: pro-active preservation is needed.***

-- UK Joint Information Systems Committee, 2002

What Gaps Exist?

Sedona Conference, 2007 (A Legal Research Consortia)

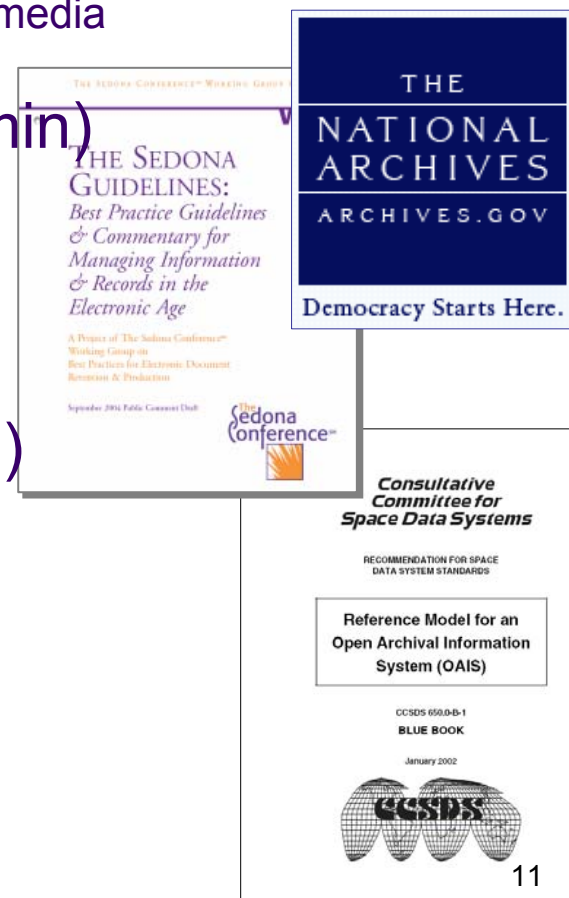
- Digital preservation requires routine **MIGRATION** to overcome software and technology obsolescence and deteriorating media

NARA (National Archives & Records Admin)

- Recommend use of current storage technology
 - ◆ ...if on disk, **MIGRATE** every 3 years
 - ◆ ...if on tape, **MIGRATE** every 5 years

OAIS (Open Archival Information System)

- No matter how well an OAIS maintains its current holdings, it will eventually **NEED TO MIGRATE** much of its holdings to different media and/or to a different hardware or software environment to keep them accessible.





Education

Requirements for Long-Term Digital Information Retention



Requirements Survey Objectives



- **Four month in-depth survey**
 - **276 respondents**
 - **IT, RIM, Legal, Security, Busn.**
 - **World-Wide, many industries**
- Designed to study long-term digital information retention requirements in the data center
- Goal: Use these requirements to frame the definition of best practices and technology solutions

Identify Requirements

Key Data Points



- Over 80% report a need to retain information over 50 years, and 68% report a need of over 100 years
- Long-term generally means longer than 10 to 15 years
- Over 40% of respondents are keeping email records over 10 years
- Database information was considered most at risk of loss
- 70% of respondents say they are 'highly dissatisfied' with their ability to read their retained information in 50 years

Key Data Points



- Current practices are too manual, too prone to error and too costly
- Collaboration is recognized as necessary in order to define information retention requirements

“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)

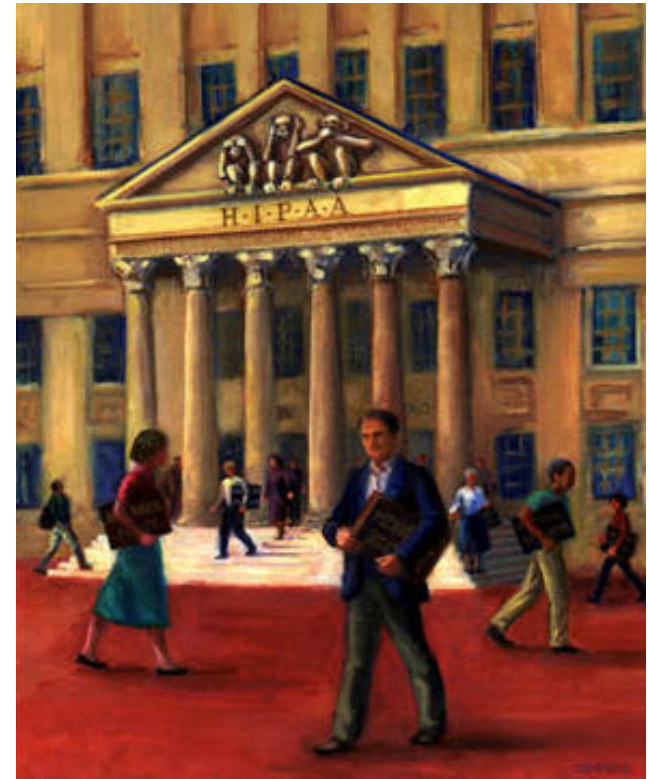
- Logical and physical migration do not scale cost-effectively
 - ◆ Only operating standard today is to migrate information physically (to new media) every three to five years and logically (to new formats) before the applications and readers die and become obsolete (every 5-10 years)
 - A never ending,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

- 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. Only 20% claimed they were confident in their ability to logically migrate some of the data.
 - › Migration will be done under crisis conditions
 - › Information is at risk long-term!

Solution Requirements



- Satisfy Business Drivers
- Overcome Inhibitors & Barriers
- Operating Requirements
- Technical Requirements

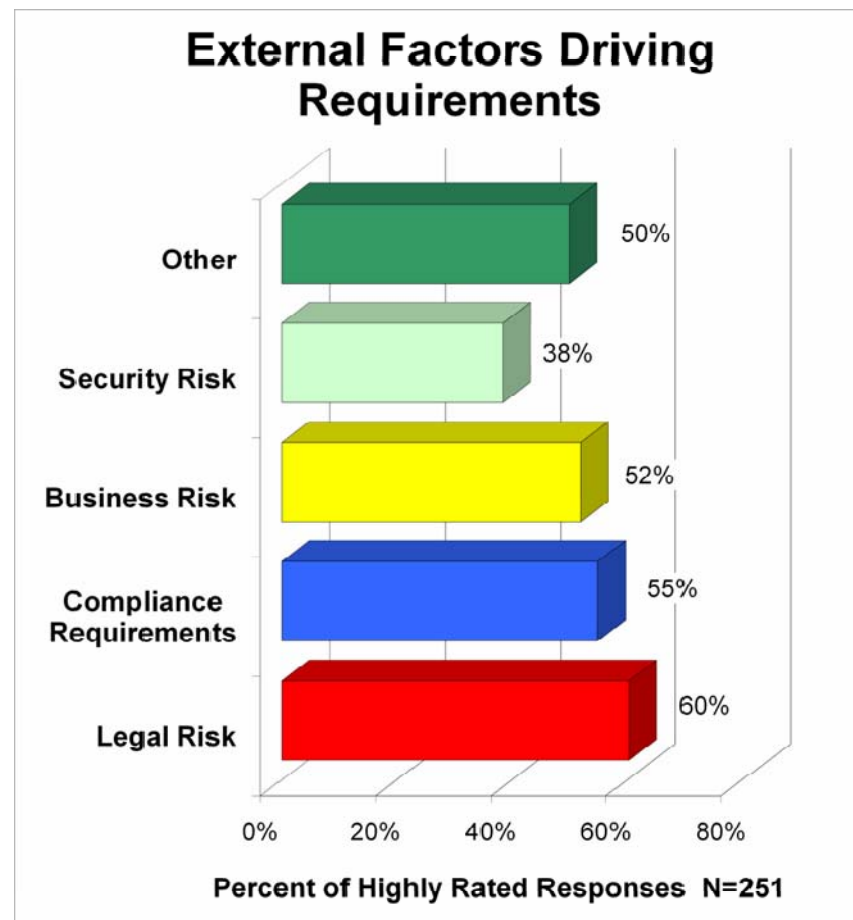


Business Requirements

Solutions must reduce

- **Legal Risk**
- **Compliance Risk**
- **Business Risk**
- **Security Risk**
- **Other**

Primarily the need to preserve the business or organization's history

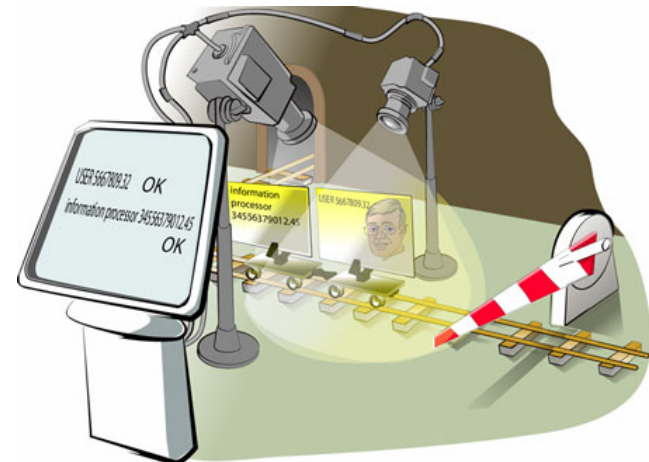


Overcome Inhibitors



- Create Value in the archives
- Reduce Cost and Complexity
- Commitment
 - ◆ Budget
 - ◆ Authority
 - ◆ Top Management Buy-in
- Elevate Professional recognition and status

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: Respondent)



Operations Requirements



➤ Set Requirements

- ◆ Collaboration
- ◆ Information Classification

➤ Use Best Practices

- ◆ OAIS, Sedona, others...
- ◆ Experience
- ◆ Clear, audited policies
- ◆ Reduce operating costs
- ◆ Better management tools

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.

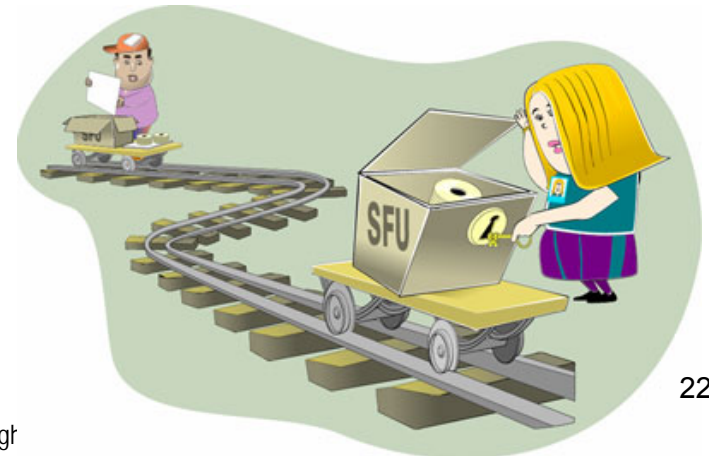


Technology Requirements



- Solve scaling and cost problems
- Standards-based Interfaces
 - ◆ Applications access information independent of storage platforms
 - ◆ Encapsulated data and metadata
 - ◆ Applications write in standard archive-formats
 - ◆ Readers for long-term data access

- Solve Migration
 - Policy based, self maintaining
 - Can not require migration projects
- Supporting Services
 - ◆ ILM-based practices
 - ◆ Discovery, Security,



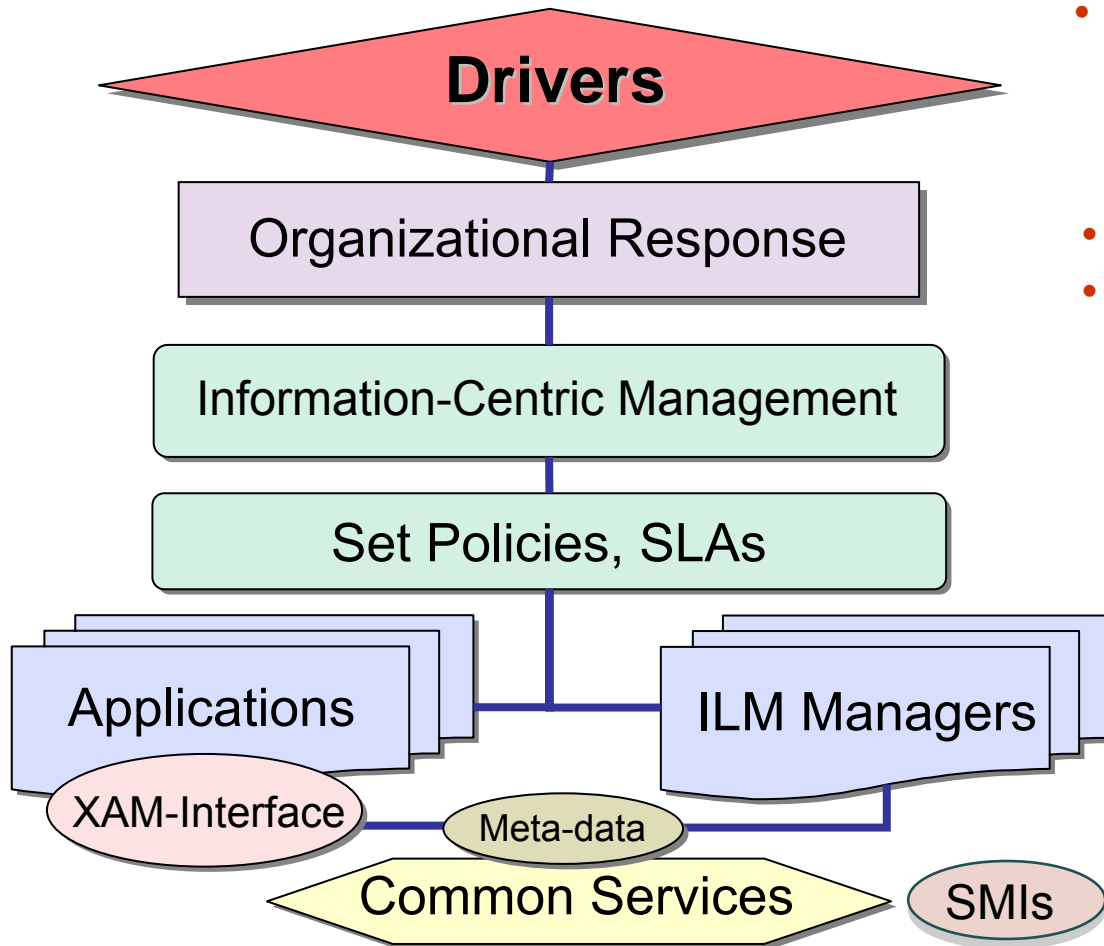
- An information and records management policy should identify and prescribe time periods for the retention of information and records that are **APPROPRIATE TO AN ORGANIZATION'S NEEDS AND LEGAL RESPONSIBILITIES**.
- There is general consensus that regardless of the various capabilities of different **BACKUP SYSTEMS**, those systems are designed for the purpose of business continuity and **SHOULD NOT BE USED** as a substitute for records management.
- What must be stored in order to achieve this goal and the manner and length of storage time will generally be decided by an organization's information technology professionals (**WITH SUBSTANTIVE INPUT FROM THE OTHER DISCIPLINES** — operational, records management and legal) as the individuals who will be relied on to manage the recovery.
- The responsible handling of electronic information and records should be considered **A CORE VALUE** of an organization.



Education

Meeting the Requirements





- **Information Convergence**
 - Compliance
 - Legal Discovery
 - Security Risk
- **Information-Centric Enterprise Collaborate**
 - Business
 - Legal
 - Finance
 - I.T.
 - RIM
 - Security
- **Identify, Classify**
- **Set requirements**
- **ILM Practices**
 - Set SLAs
 - Automate Services
 - Information
 - Data
 - Security
 - Storage

Electronic Archive:

- (Electronic) Archives are repositories for the long term storage (**meaning logical and physical preservation**) of records (**information**).

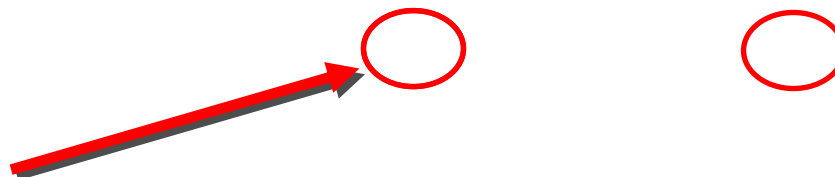
Electronic archives preserve the content, prevent or track alterations and control access to electronic records.

Sedona Conference, 2004

OAIS

Open Information
Archive System:
A Functional Model
for Electronic
Archives

SIP – Supplier Info Packet
AIP – Archive Info Packet
DIP – Display Info Packet



FEATURES

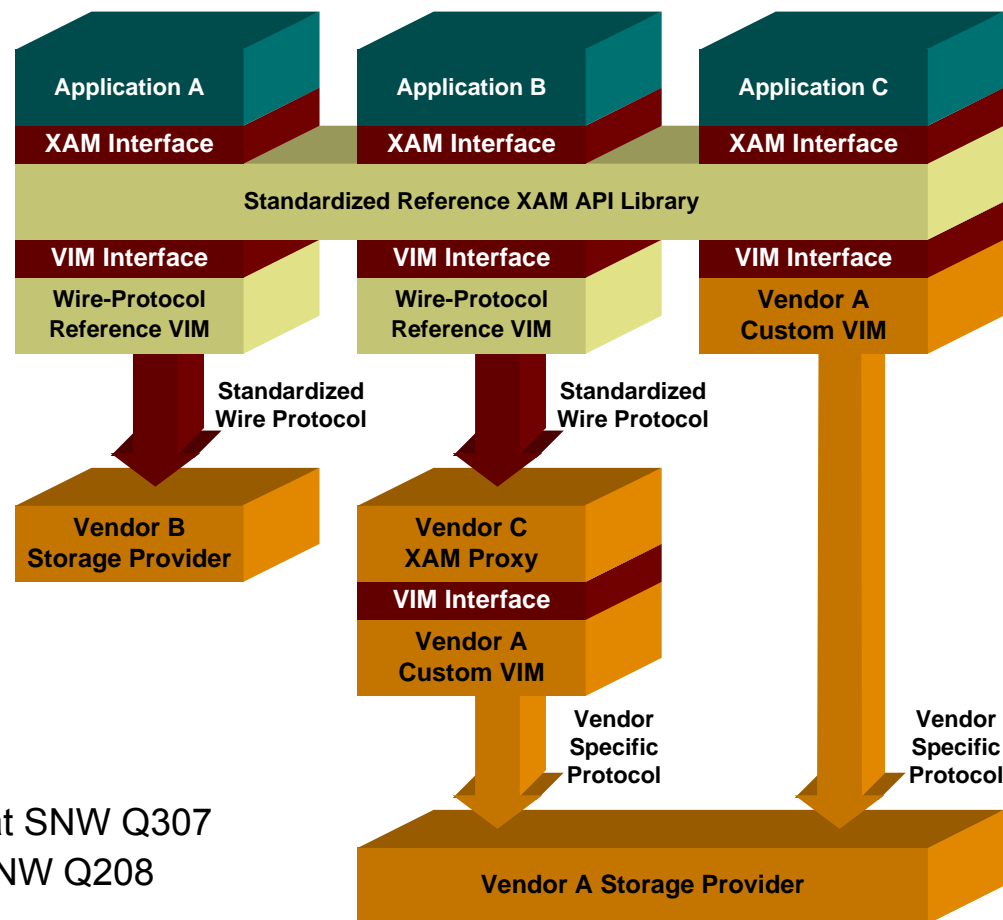
- Data independent of the application and application independent of the storage
- Standard method for accessing data (structured, unstructured & semi-structured) in object repositories that include data and metadata
- Allow data movement based on Policy without regard to storage
- Allow Information Mgmt interoperability across storage infrastructure

Timeline:

- Version #1 Specification Public Review Q207
- Early XAM product demos (vender specific) at SNW Q307
- XAM Compliant product demonstrations at SNW Q208

October, 2007

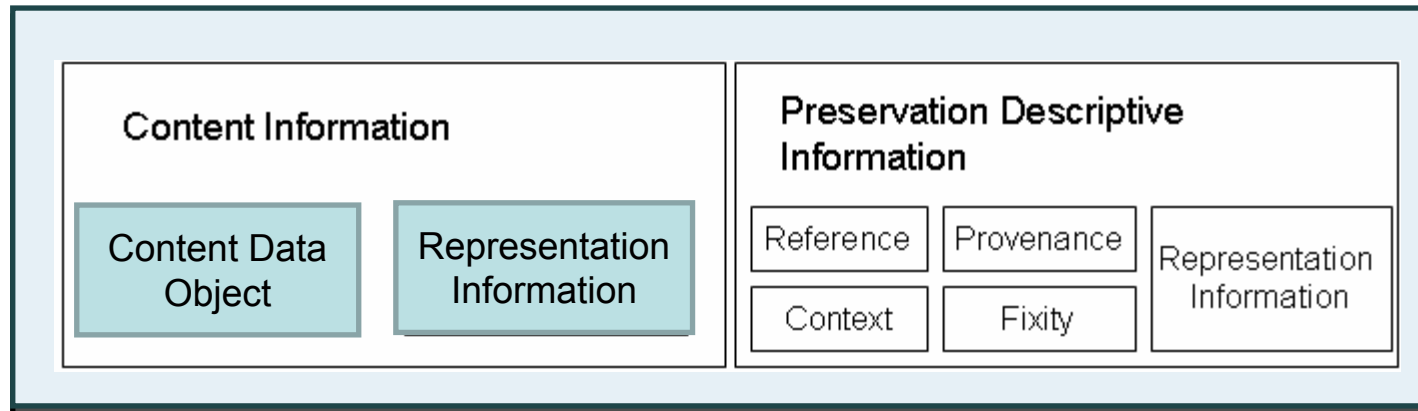
XAM – eXtensible Access Method



Standard Logical Container

OAIS Archive Information Packet (AIP)

AIP

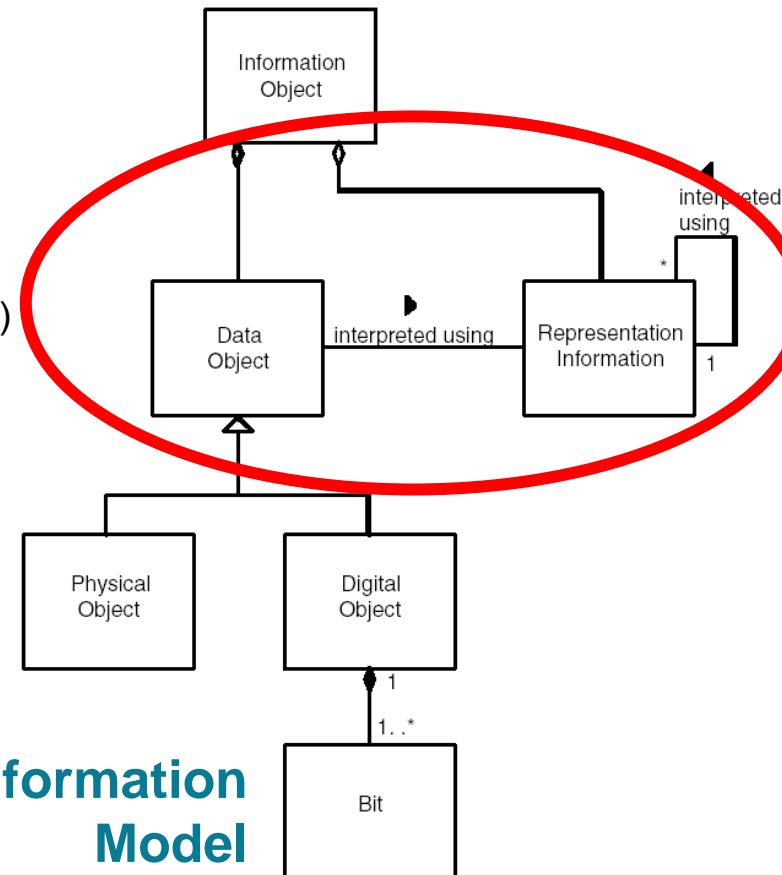


Content Object Data	the raw data that is the focus of the preservation
Representation Info	information required to interpret raw data to designated community
Reference	globally unique and persistent identifiers for the content information
Providence	history and origin of the content including chain of custody and updates
Context	purpose of content and relationship to its environment (links, etc.)
Fixity	proof that content has not been altered in an undocumented manner

Standard Logical Container

- ▶ Media agnostic
 - ◆ Tape, disk, future media
- ▶ Vendor and Platform agnostic
- ▶ Self-describing
- ▶ Support self-contained data
 - ◆ Allow representation of internal links and cross refs
- ▶ Performance
 - ◆ Good performance for large data sets (text and binaries)
 - ◆ Enable parallel reads and writes
- ▶ Interoperability
 - ◆ migrate data between systems without data loss
 - ◆ Can be interpreted in the future
- ▶ Extensible
 - ◆ Additional information which may be added in the future
 - ◆ Vendor specific extensions
- ▶ Cost
 - ◆ Free parsers
- ▶ Support additional functions on the data
 - ◆ compression, encryption, cryptography

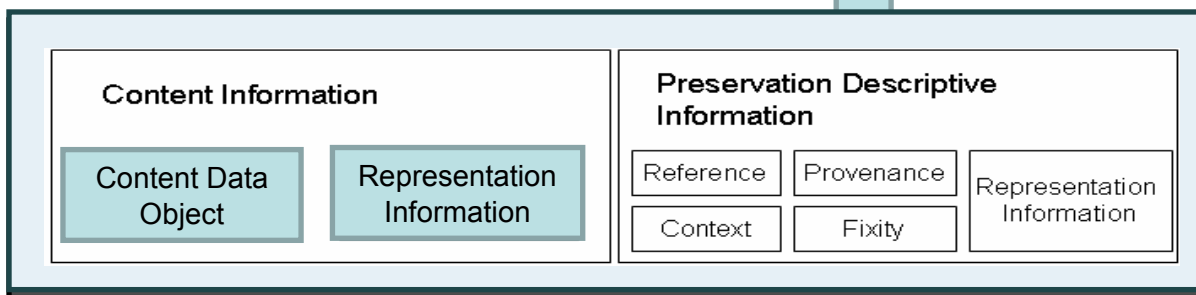
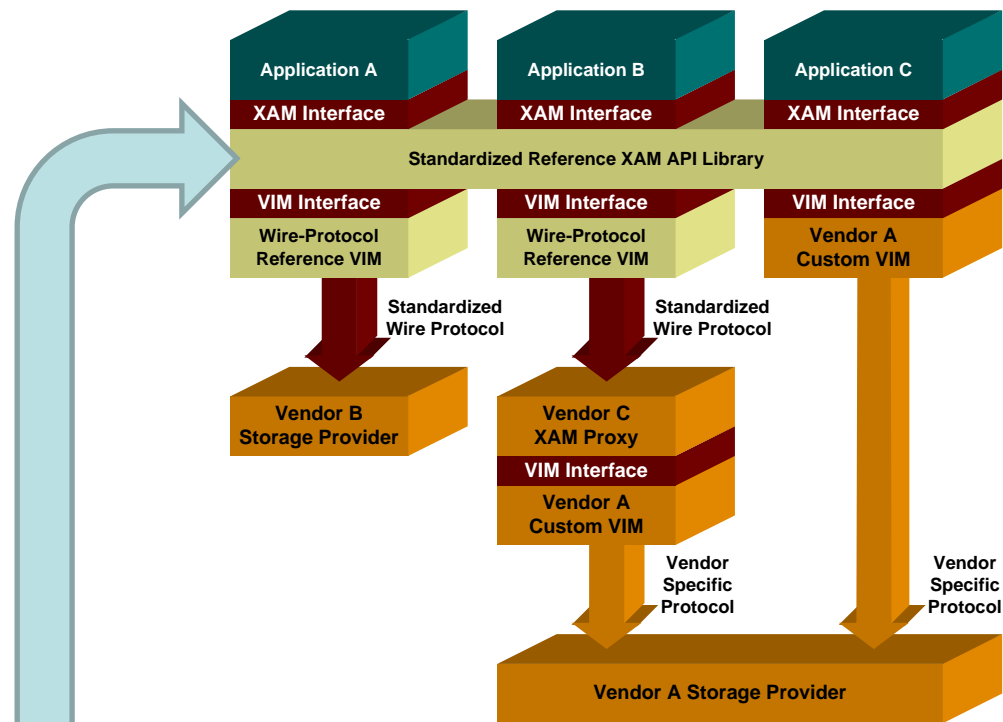
Archival Information Package



OAIS Information Model

Address Logical Migration

- **XAM Library includes a SD-SCDF container**
 - **Self-Describing, Self Contained Data Format**
- **Applications can now write a standard long-term interchangeable format**



October, 2007

Long-Term Digital Information Storage Repository

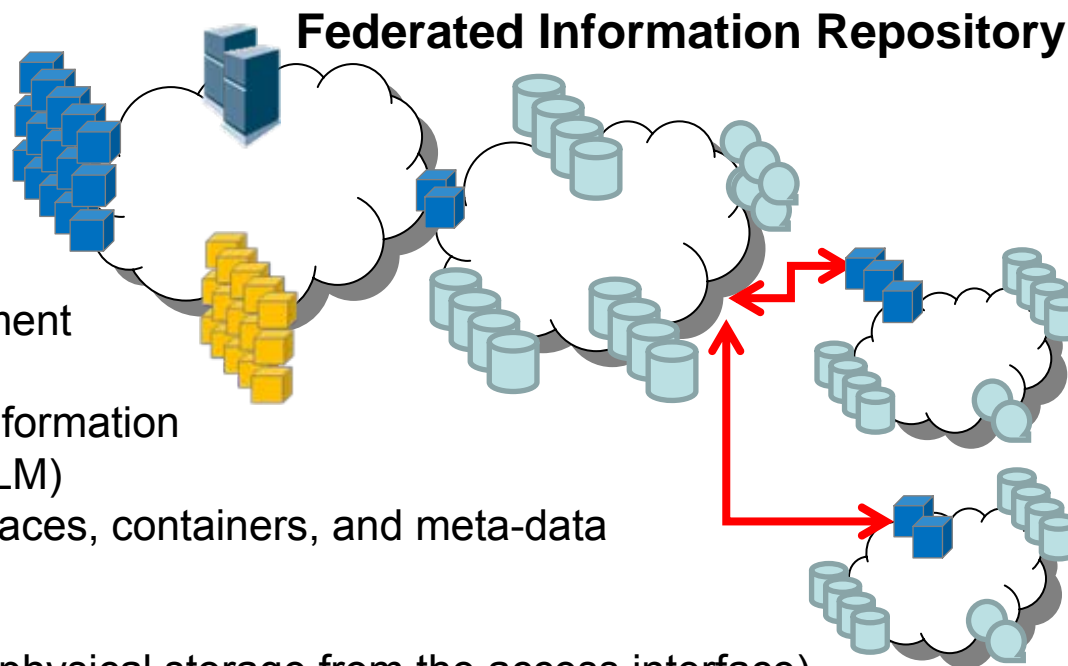
- “Migration Ready”

Federated Storage:

- Many-to-many with central management
- More than a virtualized storage pool
- Requires policy-driven, automated information management functions & services (ILM)
- Leverages XAM for application interfaces, containers, and meta-data

Features and Benefits

- Virtualized (tiered) storage (isolate physical storage from the access interface)
- Make physical migration transparent via self healing systems
- Centralized ILM-based management
- ILM-based placement, retention, deletion, protection, and business continuity
- Integrity – with CAS or WORM-like practices
- Scalable - Transparent growth with de-duplication (unique object naming schemas)
- Security – Maintains access control and access logs over time (auditable)



➤ Long-Term Retention Reference Model

- ◆ Best Practices for Storage
- ◆ Requirements (done)
- ◆ Glossary (in review)

➤ Physical Migration

- ◆ A virtualized, federated information repository in which self-healing eliminates need for special physical migration projects
 - › Add all required services (de-duplication, hash-based unique naming, location independence, encryption ...)
- ◆ Meta-data provided thru XAM

➤ Logical Migration

- ◆ SD-SCDF: self-describing, self-contained data format standard
- ◆ a container based on OAIS¹ Archival Information Package integrated with XAM
 - › Through XAM, applications can write archival formats containing metadata, authentication information, source data, and a reader.
 - › XAM encourages application support

1. OAIS: Open Archival Information System

Resources



- ◆ SNIA Data Management Forum
www.snia-dmf.org
- ◆ 100 Yr Archive Task Force
www.snia-dmf.org/100year
- ◆ NARA:
www.archives.gov/era/presentations
- ◆ OAIS – Open Archival Info System
 - ◆ ISO Reference model
 - ◆ www.public.ccsds.org/publications/RefModel.aspx
- ◆ Sedona Conference
 - ◆ www.thesedonaconference.org



Download Survey at:
<http://www.snia-dmf.org/register>

- Please send any questions or comments on this presentation to SNIA:
trackdatamgmt@snia.org

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

Michael Peterson

President, Strategic Research Corp.
Chief Strategy Advocate, SNIA Data Management Forum

Jeff K. Porter

Senior Technologist, IMSG CTO Office, EMC
Chair, SNIA Data Management Forum

100 Year Archive Task Force

Gary Zasman, NetApp
Peter Mojica, AXS-One

Get Involved !

- Find a passion
- Join a committee
- Gain knowledge & influence
- Make a difference

www.snia-dmf.org

www.snia-dmf.org/100year