Long Term Record Retention and XAM

Wayne M. Adams
Chair Emeritus, SNIA

www.snia.org

Agenda

SNIA
Advancing storage & information technology

- **Market Trends and Drivers**
- SNIA Survey
- SNIA XAM Standard
- SNIA Meta Data Work
- Summary

Information Challenge

SNIA
Advancing storage & information technology

- Escalating Storage Costs
  - Data: 70% annual increase in data volumes
  - Digital Proliferation: 92% of information is digitally created with only 30% being repurposed
  - Cost: 80% of IT budget is consumed by maintenance

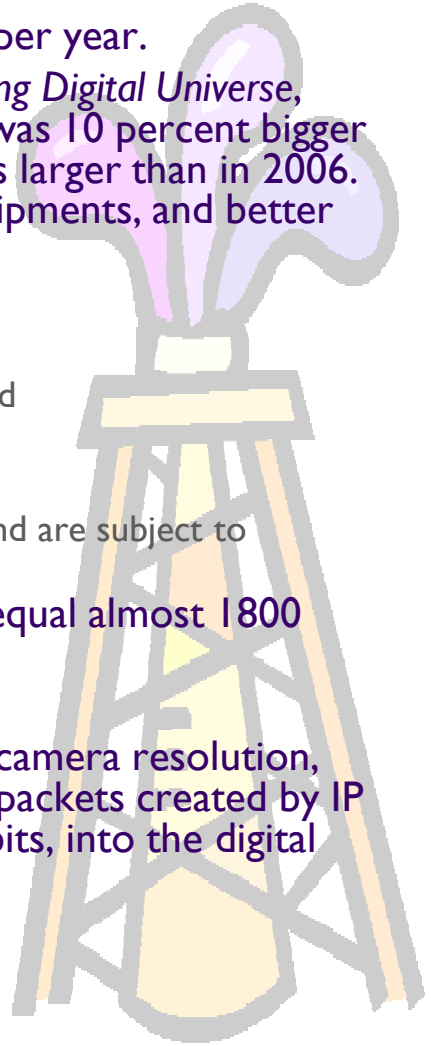
- Increasing Scrutiny and Risk
  - Risk: 80% of information is unstructured with no organizational control
  - Litigation: Nearly 90% of U.S.A. companies engaged in some type of litigation
  - Fines: $ Millions in fines for inadequate record keeping

Sources: IDC, Gartner Group, AIIM, Meta Group

Data Growth and More Growth….

SNIA
Advancing storage & information technology

- Many of endusers report information growth rates of 50 to 80 percent per year.
- According to the March 2008 IDC White Paper, *The Diverse and Exploding Digital Universe*, the digital universe in 2007—at 281 exabytes or 281 billion gigabytes—was 10 percent bigger than previously thought. By 2011, the "Digital Universe" will be 10 times larger than in 2006. The resizing comes as a result of faster growth in cameras, digital TV shipments, and better understanding of information replication.
- More findings from this study are:
  - 70 percent increase in data volumes
  - 92 percent of information is digitally created with only 30 percent repurposed
  - 80 percent of IT budget is consumed by maintenance
  - 80 percent of information is unstructured with no organizational control
  - Nearly 90 percent of U.S. companies are engaged in some sort of litigation and are subject to millions of dollars in fines for inadequate record-keeping
- In 2011, the amount of digital information produced in the year should equal almost 1800 exabytes, or 10 times that produced in 2006.
- Over 95 percent of the digital universe is "unstructured data."
- Even while image files grow to multi-megabyte size as a result of better camera resolution, the exponential growth of sensors, RFID (radio frequency ID) tags, and packets created by IP voice phone calls is streaming trillions of smaller signals, some just 128 bits, into the digital universe.



Information-Archiving Challenges: IT Perspective

SNIA
Advancing storage & information technology

Availability

I need the information when I need it

Backup

My backup challenges are as critical, if not more so, as my archive needs

Governance and Compliance

More work—big liability if not addressed properly

Consolidation

I want one solution for all of my archiving requirements… but the vendors don't offer it

I need to drive down costs

Disaster Recovery

I need to protect these critical assets just like everything else

Costs

My information is growing at greater than 50 percent per year, but I can't afford a 50 percent increase in costs

A New-Kind of Old Data
Fixed Content

SNIA
Advancing storage & information technology

- Music and Video
- Archived Email
- Medical Imaging
- Check Processing and Imaging
- Call Center Voice Recording
- IM Chat Sessions
- Video Surveillance
- Digital Photos

What is Fixed Content?

SNIA
Advancing storage & information technology

- A type of data classification that indicates the bits are no longer changing
  - Classifying this way enables storage systems to meet the requirements of this type of data
- Most data is created "fixed"
  - Photos, videos, published/emailed documents, etc.
- 70-90% of data becomes fixed at some point
  - Even transactional data becomes fixed typically within a week
- Fixed content data is GROWING at 90% year over year

Agenda

SNIA
Advancing storage & information technology

- Market Trends and Drivers
- **SNIA Survey**
- SNIA XAM Standard
- SNIA Meta Data Work
- Summary

Key Findings

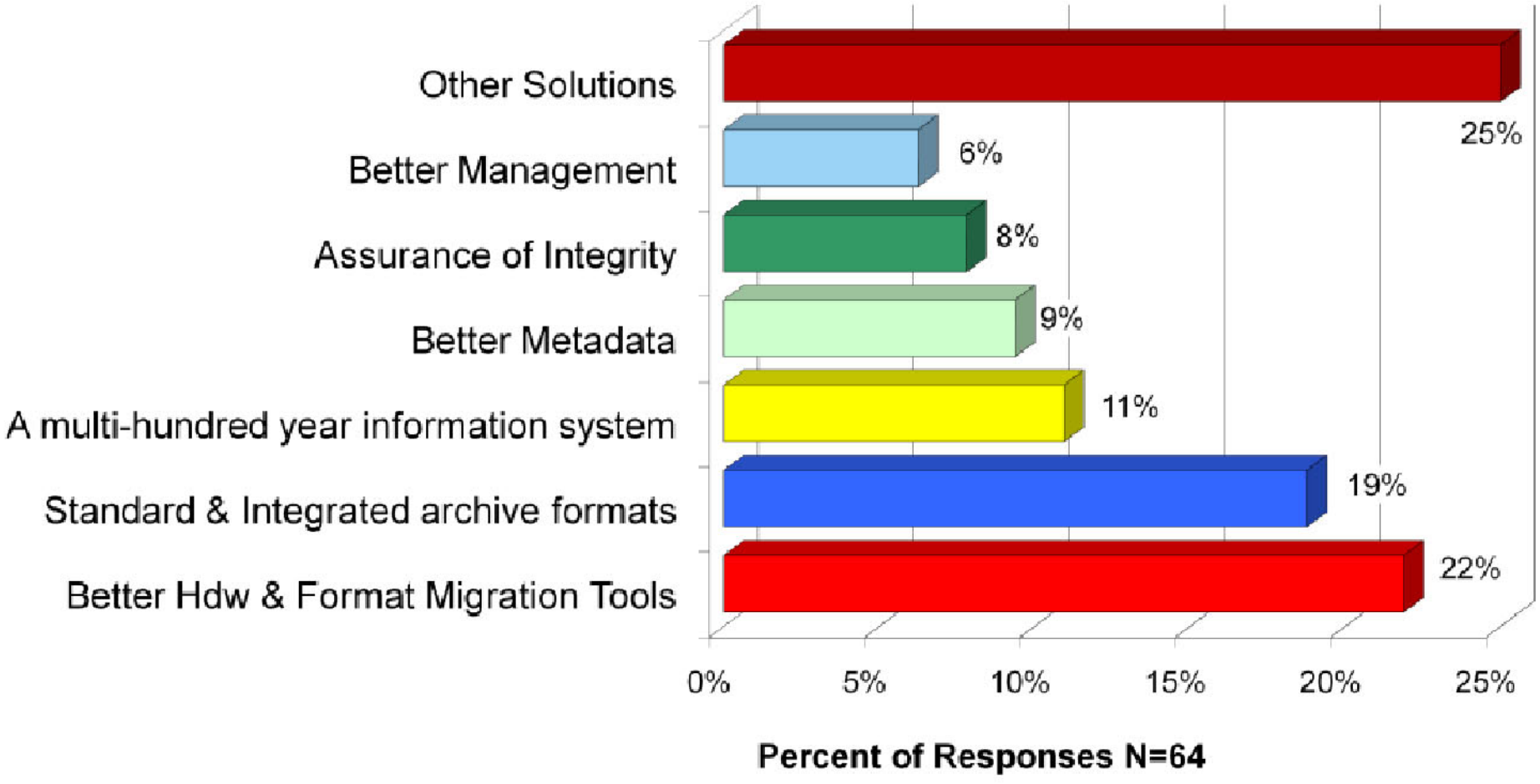
- The problems of logical and physical retention
  - Practitioners are struggling – information is at risk long-term
  - Problems are real and generally understood
- Long-term generally means over 10-15 years.
  - IT can manage to migrate and retain readability for about this long. For longer periods, 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)

SNIA 100 Year Archive Requirements Survey

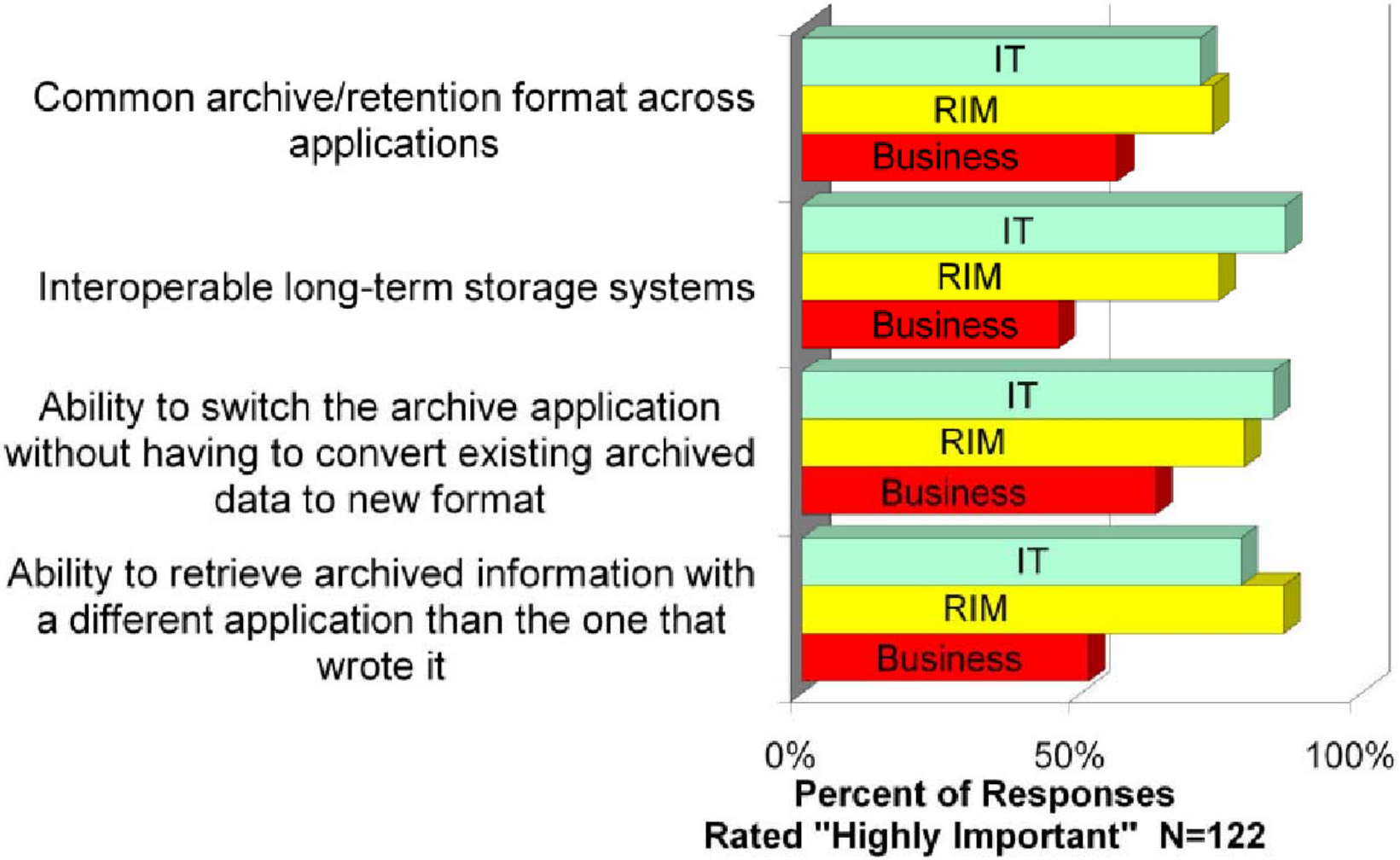
SNIA
Advancing storage & information technology

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



SNIA
Advancing storage & information technology

How Important are these Solutions to Long-Term Digital Retention?



Agenda

SNIA
Advancing storage & information technology

- Market Trends and Drivers
- SNIA Survey
- **SNIA XAM Standard**
- SNIA Meta Data Work
- Summary

What is expected of storage and storage applications?

SNIA
Advancing storage & information technology

Different groups have different expectations

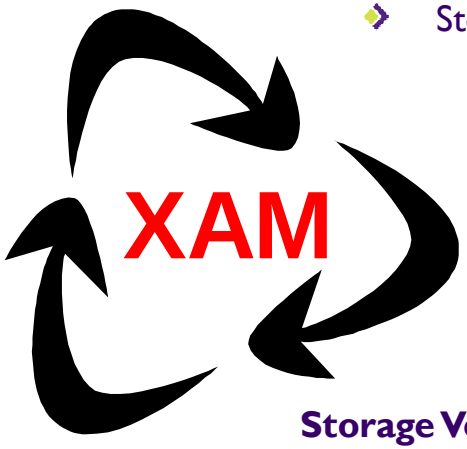
Applications Vendors

End Users

Storage Vendors

Applications Vendors

- Annotate Data with associated Metadata
- Indicate Basic Storage Management Policies
- Speak the same language to all types of devices
- Manage billions if not trillions of "records"



End Users want:

- Choices between Application Vendors
- Choices between Storage Vendors
- Easy migration between vendors/technology
- Compliance, Scalability, Performance, $/GB, Low cost of ownership (TCO)
- Store billions if not trillions of "records"

Storage Vendors want:

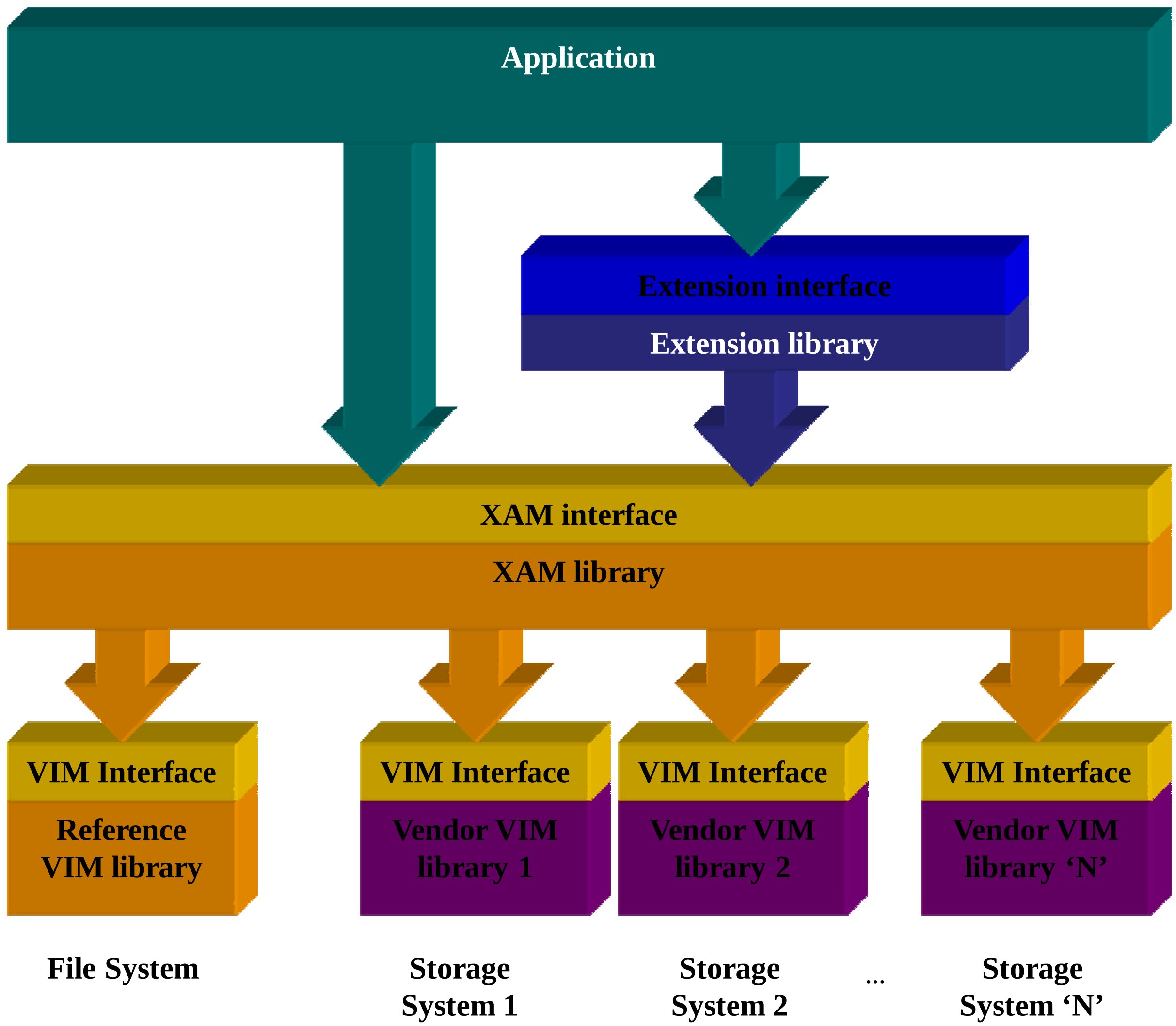
- Application Support for their Products
- Efficiently Store Application Data and Metadata
- Integrated Storage Management Capabilities
- Manage billions if not trillions of "records"

What is XAM?

SNIA
Advancing storage & information technology

- XAM is an Application Programming Interface (API) designed to address the storage needs of storage vendors, application vendors, and end users.
  - XAM was designed to enable the interoperability, storage transparency, and automation of policy driven ILM based practices for data.
  - XAM interfaces were designed with the knowledge that this data must be stored for long periods of time and with information assurance (security).
- Why an API?
  - As an API, it allows the separation of the access method from the underlying storage (block storage, file systems, or whatever)
    - This allows end users to separately choose applications and storage systems, and avoid vendor lock-in to either.
    - It allows end users to decouple storage provisioning from storage location, enabling the end user to scale their systems as required without lengthy updates to application references

What does the XAM architecture look like?



The XAM standard specifies:

- A standard library containing the API for applications
- An interface for vendors to provide plug-ins to the standard library. These plug-ins are called Vendor Interface Modules (VIMs)
- A reference VIM that allows applications to exercise the XAM functionality.

Vendors provide:

- A VIM for each unique storage system
- Vendors can also provide extension libraries, containing new functionality. Note that because these are built on top of the XAM interface, they are usable on all XAM storage systems

XAM Features You Should Care About

SNIA
Advancing storage & information technology

- XAM supports interoperability between multi-vendor storage devices
- XAM enables retention management & regulatory compliance support
- XAM enables object search and e-Discovery through extensive metadata capabilities
- XAM supports security policies
  - Access rights to the XAM object are contained inside the XAM object
- XAM provides object location independence
  - XAM data is location independent & XAM identifiers are unique across all XAM-conformant storage systems

XAM Enables Compliance & Retention Management

SNIA
Advancing storage & information technology

Regulatory Compliance

- XAM supports tamper proof storage
- Regulatory retention information (retention, disposition) is embedded inside XAM objects

Retention Management

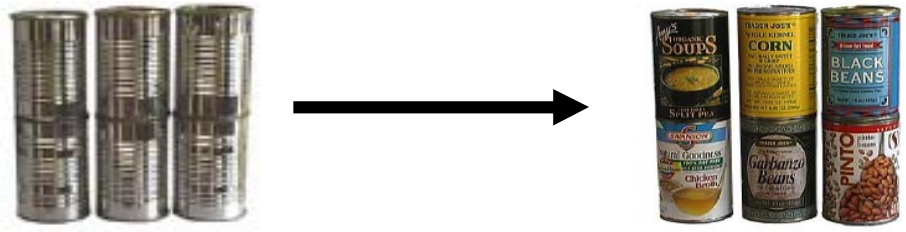
- Removes the IT management effort & cost overhead of managing retention of objects
- Reduces regulatory risk of keeping information too long… or not long enough

> "All large corporations must have a strategy for dealing with reactive discovery now and a long term plan for dealing with proactive information management…"
>
> Debra Logan
> Industry Analyst
> October 2007

The Importance of XAM Metadata

SNIA
Advancing storage & information technology

- An object without metadata is like a can without a label
  - The label contains critical information about the contents of the can
  - The XAM API supports an interface to search metadata fields



- Embedded metadata simplifies the storage & management of data for application vendors
- Embedded metadata & XAM search extends application capabilities for end users

Storage Vendor Independence

SNIA
Advancing storage & information technology

- Storage Interoperability
  - Applications / end users can select whatever XAM-conformant storage device(s) they prefer
  - End users can use multiple XAM-conformant storage devices simultaneously
- XAM Object Import and Export
  - XAM supports the ability to migrate data between XAM-conformant back-end storage systems
  - Prevents vendor lock-in

Use of MetaData Standards

SNIA
Advancing storage & information technology

Writes content and annotates it with metadata, in this case: *to, from, roles, subject* and *number of attachments*

Email Service

Email object stored by XAM SDK

Metadata accompanies content

Content

com.acme.email.from = “bugs bunny”
com.acme.email.from.role = “analyst”
com.acme.email.to = “daffy duck”
com.acme.email.to.role = “trader”
com.acme.email.subj = “what’s up doc?”
com.acme.email.numattach = 2
{ Email contents }
{ Attachment #1 }
{ Attachment #2 }

XAM specifies how metadata is represented, but not the actual metadata field names and values.

Further work is needed to standardize metadata names and allowed values for application domains like Email, Health, and Document Management.

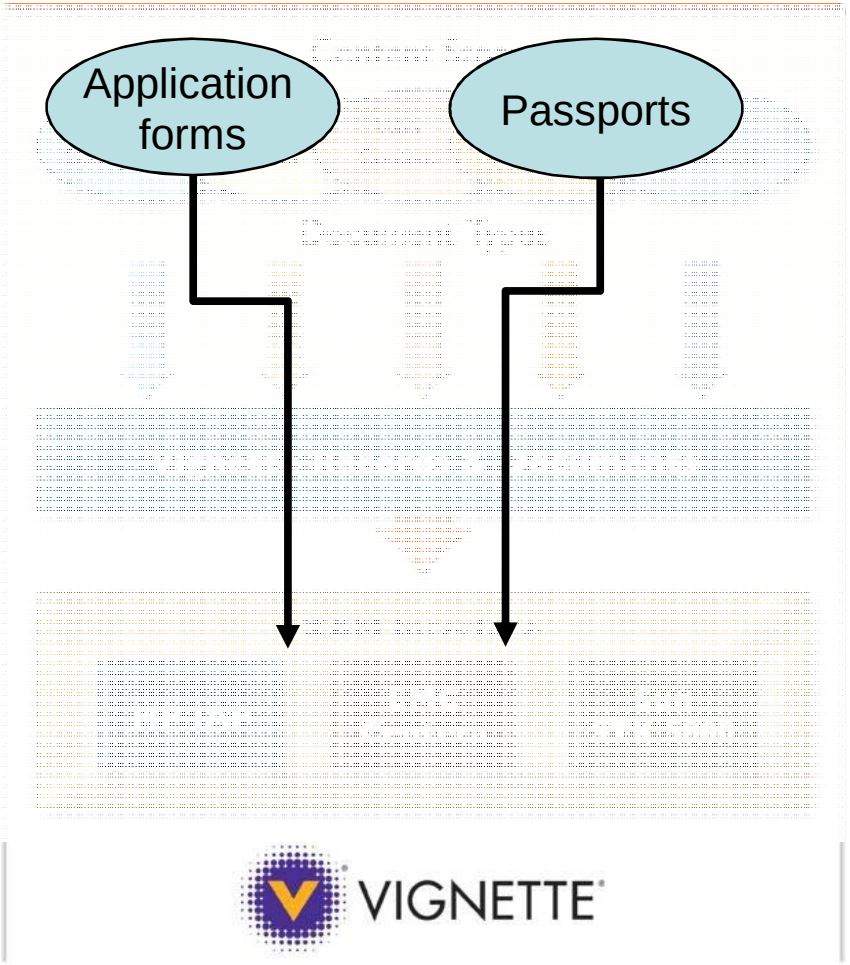
Email Analysis Program

Can access Email metadata and, without the help of the Email Service, analyze whether the sender is allowed to send to the recipient. For example, a stock analyst may not be allowed to send information to a trader.

ECM Application Use Case

SNIA
Advancing storage & information technology

- ECM Records & Documents Application port to XAM
- Full lifecycle management for enterprise content utilizing XAM
  - Application forms are routed through XAM to the Storage Platform Vendor 1
  - Passports are routed through XAM to the Storage Platform Vendor 2 and Vendor 3



"Information independence for applications and storage"
XAM makes this possible

As seen at SNW 2007 and 2008

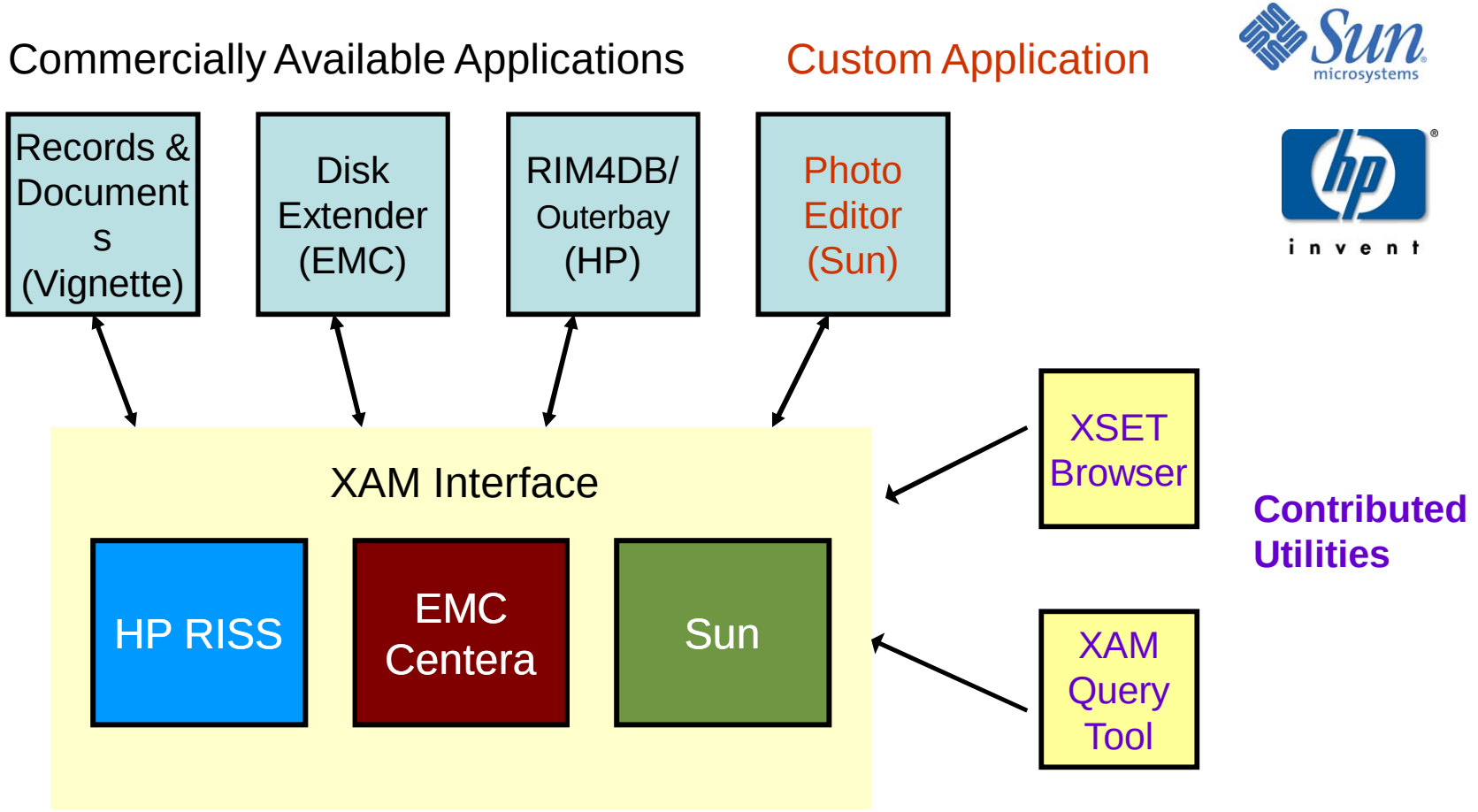
SNIA
Advancing storage & information technology

First Multi-Vendor demonstration based on XAM

EMC²
where information lives®

Sun microsystems

hp invent



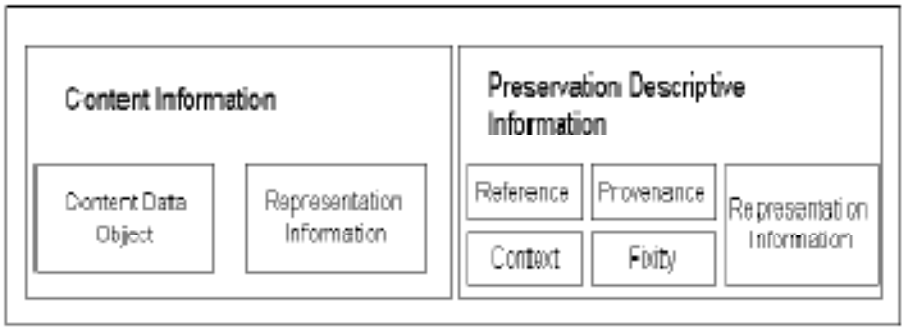
Agenda

- Market Trends and Drivers
- SNIA Survey
- SNIA XAM Standard
- **SNIA Meta Data Work**
- Summary

MetaData for Self Describing & Contained Data

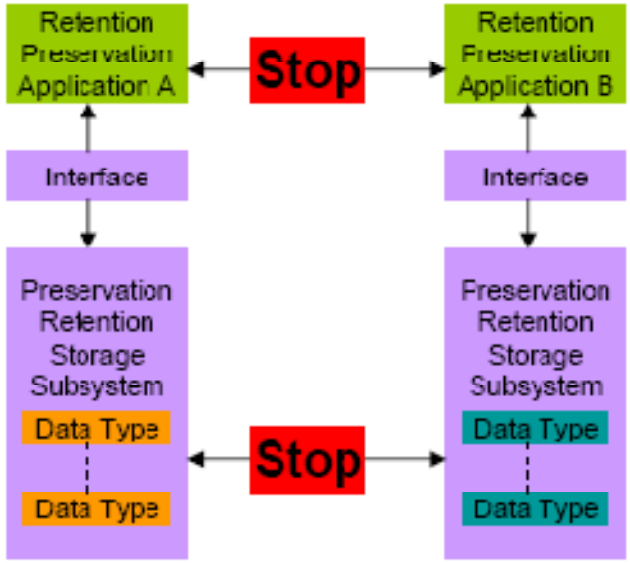
SNIA
Advancing storage & information technology

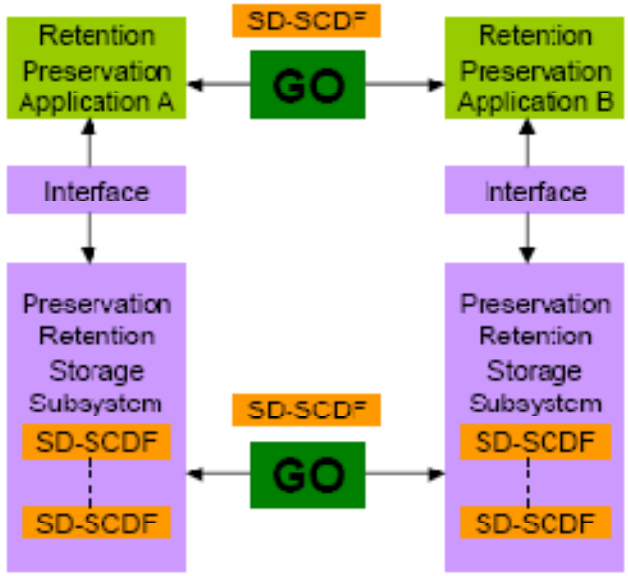
- SD-SCDF is a "container" or "wrapper" format
- Supports a number of different types of coded "content", encoded with any of a variety of applications, together with a metadata wrapper which describes the content contained within the SD-SCDF file
- SD-SCDF is being designed to address a number of problems with data formats over time, and is intended as a platform-agnostic stable standard for future archive content.
- Leverages the OAIS Archival Information Package (AIP)



Requirements to be addressed by SD-SCDF

SNIA
Advancing storage & information technology





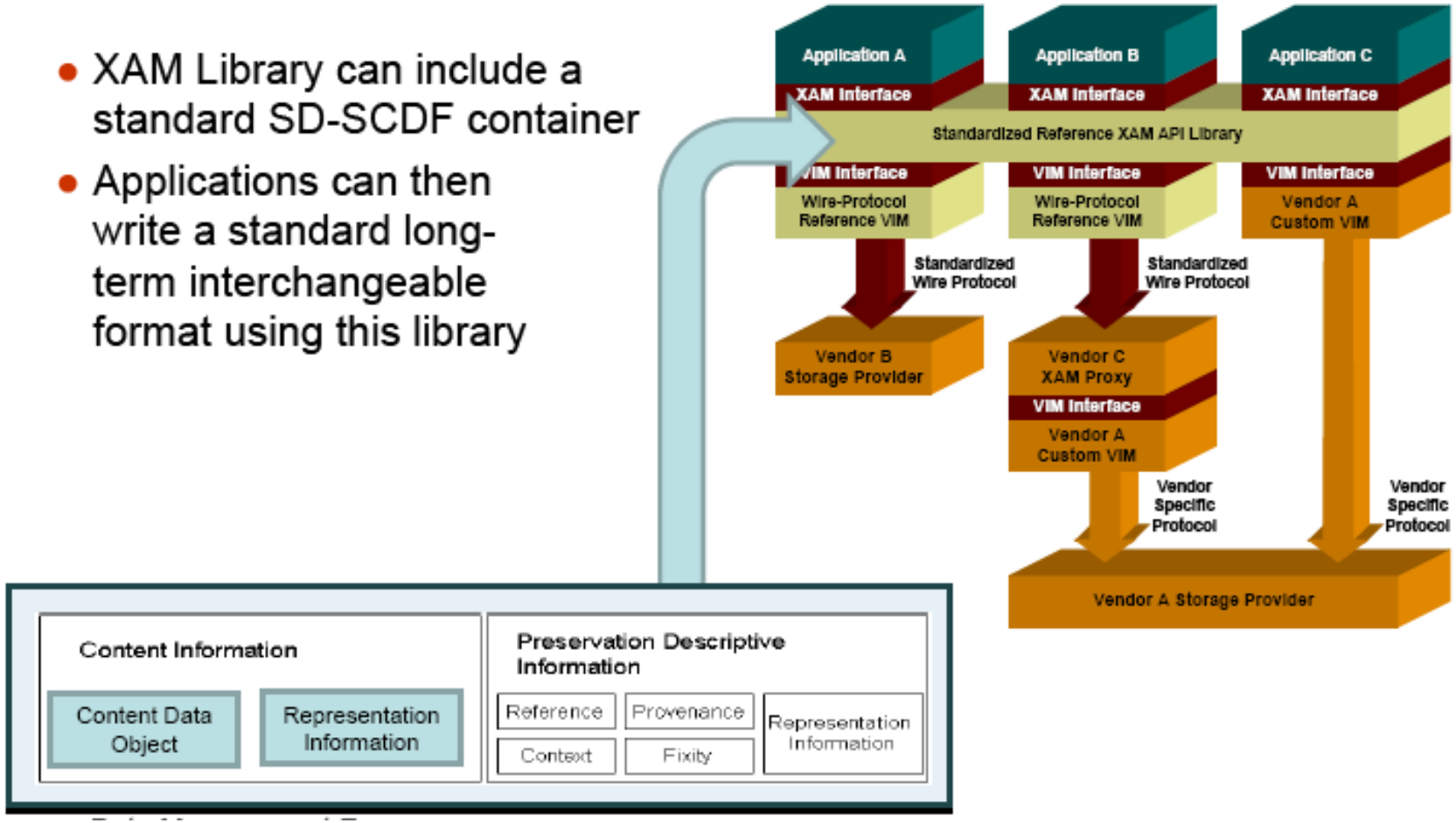
- Cannot move data and metadata between storage sub systems
- No preservation metadata standards
- Data and metadata at risk

- Can move data and metadata between systems
- Metadata supports both preservation and retention
- Supports both logical and physical migrations

MetaData for Self Describing & Contained Data

SNIA
Advancing storage & information technology

- XAM Library can include a standard SD-SCDF container
- Applications can then write a standard long-term interchangeable format using this library



Summary

- Data Growth Continues
  - Fixed Content represents one category of the data, but a large portion of growth
  - New technology and standards are desired and required to manage the growth in context of meeting business, retention, and compliance needs.
- Requirements for Fixed Content Data Management and Archiving/Archive Management overlap, but there is mutual exclusivity for each
  - Accordingly, XAM and SD-SCDF are complementary
  - Each can be used exclusively or in conjunction with each other
- Timetable for XAM
  - V1.0 Specification is now
  - V1.0 SDK is Q308
  - V1.0 commercial offerings Q408 and throughout 2009 and 2010
  - V1.0 becomes an ANSI Standard in 2009, ISO in late 09 or 2010
- Timetable for SD-SCDF
  - Draft specification – *stay tuned*

For Additional Information

SNIA
Advancing storage & information technology

- SNIA XAM Initiative
  - XAMI home page - http://www.snia.org/forums/xam/
  - XAM API specs -http://www.snia.org/forums/xam/specs
  - XAM SDK – http://www.snia.org/forums/xam/ *tbd*
  - XAM Demo - http://www.snia.org/forums/xam/flshdemo/1282_SNIA_XAM.htm
- SNIA Data Management Forum (DMF)
  - DMF home page - http://www.snia.org/forums/dmf/
  - DMF Long Term Archive - http://www.snia.org/forums/dmf/programs/ltacsi/
  - DMF 100 Year Survey - http://www.snia.org/forums/dmf/knowledge/100YrATF_Archive-Requirements-Survey_20070619.pdf
- SNIA
  - http://www.snia.org

Questions

Thank You!