SNIA Europe

SNIA / Bitkom Veranstaltung 18.Oktober 2006

Stefan Ehmann
Chairman German Committee SNIA-Europa

Archive a definition

- A collection of data that is maintained as a long-term record of a business, application, or information state. Archives are typically kept for auditing, regulatory, analysis or reference purposes rather than for application or data recovery.
- To copy or move data for purposes of retention; to create an archive(1).

SNIA EUROPE

Archiving Data

What`s the problem ?

- I have my original work-product contained in word processing documents and spreadsheets
- I saved these on an floppy disk, the 360k variety, and I need to access the data
- Today`s date is 2006
- The data was created and saved in 1986
- Just 20 years ago

Information is data that is interpreted within a context such as an application or a process

Archive vs. Backup

- Backup-Solutions are not suitable for Archiving
  - Need original Configuration to access the Data
  - No Index
  - Includes Duplicates
  - No Version History

SNIA EUROPE

Archive vs. HSM

Attributes	Archive Management	HSM
Access Method	Application can access data directly or indirectly through Archive Management	Application can access data directly. HSM is transparent to Application
Access Control	Only read access by application; application cannot modify or delete records	No limitations
Data Immutability	Guaranteed during the retention period	Not guaranteed
State of Data	Data is not in operational state but used for reference	Data is in operational state
Data Copies	Archive Management may maintain a second copy of the application data. The data archived may exist under Application control	Only one instance of data is maintained under control of the Application
Use of Tiered Storage	Can use tiered storage for cost-effective retention but not necessary	Uses tiered storage
Management Function	Manages retention, access and integrity of data, usually on tiered storage, usually set by policy	Manages transparent migration of data between tiers, usually set by policy

Long Term Archive and Compliance Indicative

- **LTACSI** is a cooperative effort of end users, IT professionals, vendors, integrators, and service providers with interests in the challenges of long-term digital information retention, archiving, and storage compliance. LTACSI is a focal point for all such activity within the Storage Networking Industry Association.

- **Our Mission**
- Enable users and practitioners in these fields to make informed choices about long-term digital information retention, archive, and regulatory compliance storage solutions. LTACSI also has responsibility to provide market development assistance for the XAM standard in development within SNIA and to support and guide other related SNIA activities.

- **Our Goals**
- Become a respected authority on both compliance and long-term digital information retention and archiving - world-wide.
- Assist end users and practitioners in understanding the challenges, best practices and available solutions associated with long-term retention, archiving, and compliance.
- Provide user requirements and market requirements in these fields to relevant SNIA activities, such as the XAM and ILM Technical Working Groups.

SNIA EUROPE

Active Projects

- Compliance Rules List white paper
- Archiving and Compliance Metrics white paper
- Fixed Content Aware Storage (FCAS) TWG charter
- The 100 Year Archive task force
- The "Business and IT Requirements for Long-Term Digital Retention" Survey

SNIA's 100 Year Archive Task

- The SNIA's 100 Year Archive Task Force is a global, multi-agency group working to define best practices and storage standards for long term digital information retention. All interested parties and organizations are invited to join in this work effort.

- **Our Goals**
- Produce with a multi-disciplinary team a "best practices for long term digital information retention" paper similar to the Sedona project
- Influence ILM as a core management and automation practice for long term archive
- Guide the impact of new storage technologies such as XAM & Grid to improve long term retention methods
- Define a standard for on-media formats & long term readability

The Problem

- Although corporate and legal issues have recently brought data archiving to the light of day, the problems associated with preserving digital information are not new. Archiving for a few years is hard enough, but when requirements dictate that data be retained for longer, problems with media deterioration and technology obsolescence can seem insurmountable…

  Source: Galen Schreck – Forrester Aug. 2005

- Long term preservation of digital content is a big challenge in the Information Society era, digital information in any form is at risk to be lost forever. Technology on which digital content relies becomes obsolete and application versions and files formats change, making data soon inaccessible. Even if content is coded in the simplest format, such as ASCII code, storage media degradation and obsolescence could make it disappear. Even on-line information such as web pages and databases, are vulnerable as much as their web structure become complex thanks to (aging) hyperlinks and cross references.

  Source: Alfredo M. Ronchi - Medici Framework

best practices today?

- *Migrate: NARA says if information is on disk drives, migrate it every 3 years and if on tape, every 5 years. But, what about the ability to read and interpret the information?*

- Physically, long-term storage is not about media life (because migration is required) but, about periodic migration to newer media. But, how do you migrate a PB every year? At some point migration becomes overwhelming.

- Logically, long-term retention is about the ability to read the information and to be able to use it. Applications have relatively short life and rarely have the ability to read information older than a few revisions. Even standard formats evolve, change, and become obsolete. At some point, information has to be migrated periodically to a new standard logical format.

- Then what about issues such as compliance, integrity, and authenticity? How are you guaranteeing these over the retention period?

XAM Standards

- The Fixed Content Aware Storage (FCAS) Technical Working Group received, reviewed and voted to accept v1.2 of the eXtensible Access Method (XAM) Interface specification contributed to SNIA in early October 2005.
- Additional information on the FCAS TWG can be found at :
  - www.snia.org/apps/org/ workgroup/fcastwg/

- The XAM interface between applications and storage systems gives applications a standard interface and metadata to communicate with object storage devices characterized as "Fixed Content Aware Storage".

- The benefits of such an interface include interoperability, storage transparency, and automation for ILM-based practices, long term records retention, and information assurance (security).

- More information is available on the public website at :
  - http://www.ltacs.org/.