WHITE PAPER

Leveraging SNIA XAM for Digital Archives

By Heidi Biggar

April, 2008

Table of Contents

New Digital Archive Challenges

As the growth of information volumes in digital archives continues to explode, retention periods lengthen and digital archives become extensions of primary production environments, organizations face a whole new set of challenges when it comes to storing, managing and protecting this data for the long haul.

From a pure capacity and performance standpoint, the list of "hows" and "ifs" can be pretty extensive. How do organizations scale their digital archives to meet future requirements? How do they provide consistent high performance? Do they add systems? If so, how do these systems communicate? What if two companies merge? How are digital archive assets (i.e., applications and systems) assimilated, or are they not? What happens if the storage systems don't support an archive application? How does interoperability and information migration occur among platforms and applications as information usage changes over time? Etc.

From a retention period perspective, just think about the number of generation and format changes that have taken place over the past 10 years; let alone the past 20+ years. Retention periods are extending beyond the lifetimes of specific media, the storage systems themselves and software applications. Today, organizations are being asked to keep various data types (e.g., mortgage documents, medical images, etc.) online for tens of years; or forever. How is the readability and authenticity of this information going to be assured in the future?

The situation is further complicated by the fact that the number of pure-play archive and archive-related products on the market has increased significantly in the past year. This is both good and bad for organizations as they implement archiving strategies. On the upside, it gives organizations the opportunity to evaluate multiple technologies (both hardware and software) and implement those that are best-suited to their environments and data requirements. Also, it drives competition among vendors, which should drive additional features/functionality into products faster while simultaneously driving down the overall costs of solutions. However, without industry standards, history has shown that this type of technology evolution often leads to interoperability issues among applications and storage platforms.

It is for this specific reason—i.e., the ability to facilitate interoperability between archive applications and storage solutions—that the Storage Network Industry Association's (SNIA's) eXtensible Access Method (XAM) Initiative was established. Going forward, IT organizations will need to be able to move archive information transparently among resources no matter where they are located and no matter how business policies change and, importantly, they will need to ensure that archive information is accessible and readable over its lifetime. With XAM as a foundation for its products, IT vendors can help future-proof archive content from the various challenges described above.

XAM and Information Access

XAM as a Foundation

The purpose of the SNIA XAM initiative is to create a standard way for disk archiving storage systems and archive applications to communicate; allowing the free-flow of information among archive devices and applications and ensuring the readability of archive information for years to come.

The XAM standard ensures that any XAM-compliant application can work with any XAM-compliant storage system, and vice versa (see SNIA Definition of XAM). It provides a standard interface (initially an API) and metadata to communicate with object storage systems. This interface allows data to be moved among XAM-compliant systems as policies dictate or hardware is swapped in/out of the environment, etc.

SNIA Definition of XAM

"The XAM-Interface is an application-level interface coordinating information metadata between applications and storage systems. This new interface (initially APIs) gives applications a standard interface and metadata to communicate with object storage devices such as those characterized as "Fixed Content Aware Storage" to achieve interoperability, storage transparency, and automation for ILM-based practices, long-term digital information retention, and information assurance (security)."

For more information, go to http://www.snia.org/forums/dmf/programs/xam/.

Why a new standard and not an existing one? The answer is simple: Fixed content, or unstructured, data has very different use, retention and access requirements than structured data types. While digital archives can easily contain millions, billions or even trillions of pieces of unstructured data, the business value of this data is different from the production data (or structured data) stored on primary storage systems. Archive data is historical in nature; production data is transactional in nature. This means fewer IT resources and dollars are typically allocated to storing, protecting and managing archive information than production data. However, to get the maximum business value out of archive information, it needs to be easy to access and manage over the long-term. XAM facilitates this process.

XAM Specifics

XAM creates a globally unique name (or address) for each fixed content object and stores associated metadata about each object. The globally unique address is not linked to any particular application (the data is de-coupled from the application), storage hierarchy or storage system. This has several important benefits:

- Archive information can be quickly accessed from any application or storage device that is XAM-capable.
- Archive information can be moved and shared transparently—according to policies—across XAM-capable applications and storage devices.
- Because the unique address is derived from the content, the authenticity of the archive content is ensured for the life of the object. If the content changes, a new address is given.

The metadata, meanwhile, is used for performance; management purposes such as searching and indexing, data movement or migration; and data classification. Think of the metadata as the label on a can of vegetables in the grocery store. These labels tell shoppers what's inside the can without opening it. Now think about how difficult it would be to find what you need in a supermarket if there were no labels on cans? Similarly, metadata, which provides pertinent information about each piece of content, allows organizations to search and manage data without having to "open up" each piece of content to see what's inside. This is different from NAS or SAN systems, which store data as files and blocks, respectively.

Together, the globally unique address and use of metadata provided by XAM will enable organizations to scale their archive, access content independent of the application that stored it and easily manage archive repositories.

XAM Goals

SNIA has established some clear objectives for XAM. These include:

- Advancing the XAM specification to first become a SNIA Architecture, then an ANSI standard and later an ISO standard.
- Ensuring the adoption, implementation and deployment of the XAM programming interface(s) by applications concerned with fixed content data management and storage systems, including SAN, NAS and CAS.

- To build and serve a XAM community that includes storage vendors, independent software vendors and end-users to ensure that the specification fulfills market needs for a fixed content data management interface standard.

EMC and XAM

SNIA did not set out to reinvent the wheel with XAM. Since EMC effectively created the category and EMC Centera has the most run time field experience dealing with various archived objects, it made sense for the industry to adopt a significant amount of the work that had already been done. Emerging or turbulent markets—where competitors defend their own interfaces or connection technologies—will always try to position "their way as the right way" when it comes time to form standards. It has happened in everything from SCSI to TCP/IP to Postscript. In this case it made much more sense to simply adopt what the market was already leveraging, which just so happened to be EMC.

First introduced in 2002, EMC Centera is a purpose-built disk-based archive system that leverages a globally unique address scheme and associated metadata to store and manage fixed content written to the platform..

Beyond this, SNIA sought to leverage the experience and the knowledge of 50+ vendors to develop an interface that benefits end-users and application/storage system vendors alike. XAM provides the standard means of communicating in archive environments, leaving it to individual vendors to differentiate their products with specific features and functionality (e.g., search and indexing, data movement and migration, etc.).

EMC is a member of SNIA's Fixed-Content Aware Storage (FCAS) Technical Working Group (TWG) and XAM SDK TWG. The FCAS TWG is chartered by SNIA to develop standards around fixed-content aware storage, including the XAM interface. XAM SDK TWG is focused on developing a XAM software reference implementation and improving the consistency of implementation for better interoperability.

EMC Centera

EMC Centera's customer count now exceeds 4,500, accounting for more than 230 PB of capacity shipped to date. EMC Centera's feature set includes:

- The ability to self-configure, self-heal and self-manage; allowing organizations to easily manage (with minimal IT resources) petabytes of information.
- Single-instance storage—reducing the amount of physical capacity that is needed. The same check image or mortgage document is stored only once.
- The ability to add capacity non-disruptively.
- Assured information authenticity and online access.

Summary

As archive volumes grow, the ability to move and manage these information resources—and share this information among multiple applications and storage systems—will become increasingly important. While products that provide some level of heterogeneity on the front- and back-ends currently exist, XAM takes it to the next level.

XAM make it easier to leverage content stores, implement retention policies & access rights, and many other functions in a common way across applications and storage solutions. XAM as a standard will be the "common interface" for all types of applications and devices. It virtualizes object-based archive environments—masking the complexity of interoperability of these environments from users. XAM allows organizations to eliminate stove-piped silos of archive data and allowing users to more easily share, search and leverage archived information for IT and business benefit, which is the ultimate goal of data archiving.

With XAM capable solutions, individual storage vendors and application providers will continue to differentiate their products, as EMC has with its EMC Centera platform, with advanced feature sets.

20 Asylum Street
Milford, MA 01757
Tel: 508-482-0188
Fax: 508-482-0218

www.enterprisestrategygroup.com