

# XAM standard addresses archiving issues

By Christina Casten

**SNIA** FIXED CONTENT CONSISTS OF DATA that doesn't change over time, such as digital images, e-mail messages, presentations, video content, medical images, and check images. Fixed content data must be kept for long periods of time, often to comply with government-regulated retention periods. Some industry analysts say fixed content is growing in volume at a far faster rate than transaction-based data and could account for more than half of all corporate data within a few years.

There is no viable long-term strategy to ensure digital information will be readable in the future. Digital documents are vulnerable to loss via the decay and obsolescence of the media on which they are stored, and they become inaccessible and unreadable when the software needed to interpret them, or the hardware on which the software runs, becomes obsolete.

Historically, storage vendors have approached this problem with new technology such as content-aware storage (CAS), which is designed to store large volumes of fixed content over extended periods of time. Unlike NAS, which is designed to facilitate collaboration and file sharing, or SANs that focus on data sharing and performance, CAS is specifically designed for fixed content that might have a significantly extended lifecycle compared to transactional data.

The primary access method for CAS is via an application programming interface (API) that supports the metadata and other advanced features of CAS. Alternatively, CAS storage devices may also be accessed via more traditional methods, such as NAS protocols, FTP,

and HTTP, although at the cost of limited functionality.

Due to the proprietary nature of CAS, the Storage Networking Industry Association (SNIA) recognized that the long-term digital information preservation problem calls for a solution that does not require continual human intervention every time formats, software, hardware, document types, or record-keeping practices change. The approach must be extensible, and it must handle current and future documents of unknown types in a uniform way. This is the goal of the eXtensible Access Method (XAM) specification under development by SNIA.



CHRISTINA CASTEN

## XAM on the way

According to the Enterprise Strategy Group (ESG), fixed content is growing at an annual rate of 92%. All archives use some form of metadata for description, reuse, administration, and preservation of archived objects. XAM addresses the rapidly growing and increasing role of fixed content storage by providing applications with a standard interface and metadata to communicate with fixed content.

The XAM interface specification defines a standard access method between “consumers” (applications and management software) and “providers” (storage systems) of fixed content. XAM annotates objects with metadata, which allows policies to make intelligent decisions about the management of objects, without referring back to the application. Benefits include the following:

- XAM stores content as “objects” that consist of data and annotated metadata. An example is an X-ray image that is stored as the data component, while the patient’s name and other medical record details are stored as attached metadata. Metadata could be stored in a uniform format that can be indexed and searchable by independent tools. Also, metadata can help record and save important contextual information about the data, which may be used to interpret and make use of the data many years later;
- XAM eliminates the use of proprietary APIs; and
- XAM generates a globally unique name (address)

for each object, which is independent of the current computing environment, organization, location, or technology. As a consequence, objects may move around freely in time, changing their physical or technological location, all transparent to their current owner. This property is a fundamental enabler for transparent information lifecycle management (ILM).

## Benefits of XAM to ILM

It is expected that the emergence and adoption of XAM-based technology will have a positive impact on the adoption and implementation of ILM-based practices. XAM provides location independence for stored objects and can efficiently manage the content without the application needing the specific physical location of that content or technology on which it resides.

XAM also raises the level of importance of contextual information housed in metadata (data about the

Ultimately, end users will benefit from storage technology independence.

content being stored) to the same level of importance as the content itself. By bundling content and metadata together, applications can easily manage and share information about stored content. And, as an interface between applications and the physical store, XAM metadata allows policies to make intelligent decisions about the management of objects without referring back to the application—a primary goal in ILM.

Over time, XAM-compliant devices will be one set of the available resources for maintaining the longevity of business data. By abstracting the physical assets of storage with a data persistence perspective, the underlying storage can age and run the course of its product lifecycle while the XAM architecture provides a consistent view of the managed content and all of its attributes. This capability is important as the adaptive data

## Archiving benefits of XAM

- Applications and data will be portable among and between applications and underlying storage devices because XAM decouples the software application from the storage platform;
- An unlimited number of objects can be stored with XAM since the object is independent of the platform and not subject to limitations of file systems; and
- The archive can be searched without involving applications since metadata is bundled with objects.

## SNIA XAM Initiative

The SNIA XAM Initiative aims to drive adoption of the forthcoming eXtensible Access Method (XAM) specification.

This initiative will serve an XAM community that includes storage vendors, independent software vendors, and end users to ensure the specification fulfills market needs for a fixed-content data management interface standard. These needs include interoperability, information assurance (security), storage transparency, long-term records retention, and automation for ILM-based practices. To join the XAM Initiative, visit [www.snia.org/xami/join](http://www.snia.org/xami/join).

In addition to the XAM Initiative, the two XAM-related technical work groups are FCAS and XAM SDK.

### FCAS

The Fixed Content Aware Storage (FCAS) TWG serves as a center of technical activi-

ties related to new application-level interfaces for storage of unchanging data (fixed content) and associated metadata based on a variety of naming schemas, including CAS and global content-independent identifiers.

### XAM SDK

The XAM Software Development Kit Technical Working Group (XAM SDK TWG) is chartered to develop software that implements current and future versions of the FCAS TWG XAM specifications. This software (binaries and source) will go through the SNIA software development process and will be made available to non-SNIA members.

To join SNIA and its Technical Working Groups, visit [www.snia.org/join](http://www.snia.org/join).

center will also be able to age and retire physical assets while preserving the various services through other abstractions for the different types of resources.

In addition, based upon the SNIA Data Management Forum (DMF) 100-Year Archive Requirements Survey report ([www.snia-dmf.org/100year](http://www.snia-dmf.org/100year)), work is underway to leverage and integrate the Open Archival Information System (OAIS) ISO Standard's concept of an "archival information package" with XAM to solve the logical migration challenges of the future. The interface is intended to achieve interoperability, storage transparency, and automation for ILM practices, long-term records retention, and information assurance (security).

Ultimately, end users will benefit from storage technology independence as well as increased mobility of their data. In

addition, some of the challenges associated with migrating data from different storage devices within an electronic archive (which typically occurs every three to five years) are alleviated.

Recognizing the importance of these benefits, more than 95 individuals from 45 companies representing storage vendors, application providers, and end users are currently contributing to the XAM specification development effort. The first multi-vendor proof of concept demonstration based on XAM was provided at the Storage Networking World Fall 2007 conference in Dallas. For more information, go to [www.snia.org/forums/xam/](http://www.snia.org/forums/xam/). □

**Christina Casten** is co-chair of the Storage Networking Industry Association's XAM Initiative.

## Evaluating options for file system software continued from p. 31

also critical business intelligence that may differentiate the organization from its competitors or provide necessary information for regulatory auditors and investors. File-based data that falls into this category is stored for long periods of time and requires high availability and integrity. In such environments, there is a challenge to deliver scalability, availability, and ease of use yet keep the costs down.

A federated file system can address the challenge by leveraging existing investments (file servers or NAS appliances) through file virtualization. File virtualization software centralizes the management of existing file resources, simplifies migration, improves scalability, and simplifies overall management of the environment. However, as performance requirements increase, a federated file-system architecture may become a bottleneck.

### Replacing file server/NAS environment

Another approach to solving the increasing challenge of managing file-based data is to replace the file server/NAS envi-

ronment. The new solution may involve a different federated file-system architecture that minimizes the performance impact of a metadata node by limiting

## Enterprises are facing the challenge of identifying and selecting a file system option that fits the criteria for a given application.

the exchange between the clients and the master node to metadata queries. Since only a small amount of data has to be exchanged, the impact on performance is minimized. Another approach is a clustered file system that delivers parallel processing capabilities and scales with the addition of every new server node. In this architecture, files may be moved to different tiers of storage without impacting end users.

File-system software is becoming an important component of the data-center infrastructure. This is especial-

ly true for organizations that rely on storing and delivering file-based data as their primary revenue-generating activity. These include sites hosting pho-

tographs, providing streaming videos, and delivering online storage space for personal data.

As data-center managers strive to tame the explosion in file-based data, they look to file system software developers to deliver value-added features that can help not only manage, scale, and protect the data, but also help minimize its impact on data-center operations. Features such as automatic migration, storage tiers, storage optimization (e.g., compression and de-duplication), and security can make a significant dif-

ference in power and cooling consumption, data retention and protection, and personnel resources required to manage the environment.

### Conclusion

Obviously, all IT requirements are not equal, and one-solution-fits-all no longer applies. The diverse requirements of different market segments call for the development of solutions that excel at addressing each organization's specific issues. Enterprises are facing the challenge of identifying and selecting a file system option that fits the criteria for a given application. To make the right decision, they must first understand the options. □

**Noemi Greyzdorf** is a research manager at International Data Corp. (IDC).

### BULK REPRINTS

Do you need bulk reprints of an *InfoStor* article?

If so, contact **Andy Speter**  
Tel: (212) 221-9595 x250  
Fax: (212) 221-9195  
[andy@parsintl.com](mailto:andy@parsintl.com)