



# Tech OnTap

## Highlights

May 2008

- › [100 Year Archival](#)
- › [NetApp Accelerate](#)
- › [Exchange 2007 Best Practices](#)
- › [Premium AutoSupport Case Study](#)
- › [Tech OnTap Archive](#)

## Top Three Best Practices for Exchange 2007 Replication

Robert Quimbey, Microsoft Alliance Engineer



Are you getting the most from Exchange 2007? Unlike past service packs, Exchange 2007 SP1 includes dozens of new features and many important fixes for replication.

Check out this best practices guide for replication in NetApp® environments and find out how you can increase availability, decrease I/O workload, and improve transport performance.

› [More](#)

## Tips from the Trenches

### 100 Year Archival: Are You Ready?

Gary Zasman, Global Practice Director



In a recent SNIA survey, 68% said they had some data they needed to retain for 100 years or more. Find out what it will take to save data for the long haul.

› [More](#)

## Blogging with Dave



Dave Hitz  
NetApp  
Founder  
and EVP

### Server Virtualization: Just Starting or Almost Done?

Given the history of the industry, it seems unlikely that server virtualization will drive costs down in the way people think.

› [Dave's Blog](#)

## Get Involved With NetApp

### Download a Full Set of NetApp Visio Stencils



Check out these photorealistic physical front and rear views, 3D isometric topology shapes, presentation icons, and more.

› [More](#)

## Tech OnTap Spotlight

### › [NetApp Accelerate Poll: How Can We Rock Your World?](#)

Tell us what type of entertainment you would like to see at NetApp's first annual user conference.

### › [Premium AutoSupport Case Study](#)

See how Alliance & Leicester uses health check, event alert history, and upgrade planning.



# Tech OnTap

## Quick Links

- › [netapp.com](http://netapp.com)
- › [Tech OnTap Archive](#)

## Getting the Most from Exchange 2007 Replication Best Practices for SP1

By [Robert Quimbey](#)

About a year ago, I wrote [an article that appeared here in Tech OnTap](#) discussing, among other things, the new replication capabilities in Microsoft® Exchange Server 2007. The recent release of Exchange 2007 Service Pack 1 (SP1) not only provides bug fixes and enhancements to the capabilities I described, it also includes substantial new functionality, and it affects the way you should design the architecture of your NetApp® storage for Exchange. This article is an overview of the changes that have the greatest impact on storage and a reexamination of best practices for designing Exchange storage. For details on these topics, see my [recent technical report, "Microsoft Exchange 2007 SP1 Continuous Replication Best Practices Guide."](#)

### Important Enhancements and New Features

Among the many changes in Exchange 2007 SP1, there are five that you should definitely know about.

**SP1 is the only version of Exchange 2007 that supports Windows® Server 2008.** Windows Server 2003 is still supported under SP1 as well.

**SP1 offers a new flavor of replication, Standby Continuous Replication—SCR.** SCR lets you separate high availability and site resilience functions and lets you create a lot of new redundancy scenarios. It can be configured for use with any type of Exchange server as long as it's not also configured for Local Continuous Replication. (You can't be using LCR on your target SCR servers either.) You can have multiple SCR targets for each storage group. For instance, you might have both a local and a remote replica. You can also configure a single SCR server as a target for multiple source servers.

**SP1 completely redesigns how storage is used on replication targets to make it much more efficient.** The I/O on all replication targets (LCR, CCR, and SCR)

## Explore

### More About Exchange 2007

If you've been thinking about Exchange 2007, maybe it's time to find out more. An earlier Tech OnTap article by the same author answered several important questions:

- How does Exchange 2007 affect my backup strategy?
- How does Exchange 2007 affect my disaster recovery strategy?
- Which storage networking protocol should I use?

› [More](#)

### Implementing Exchange Archiving

The term "e-mail archiving" can be used to refer to archiving for mailbox management needs or journaling for regulatory compliance, but these are very different problems with different solutions.

A recent Tech OnTap case study explains the difference, describing how one financial services company:

- Migrated most users to Exchange Server 2007
- Added archival capability to get the volume of e-mail under control and eliminate the need for PST files
- Added journaling to meet regulatory requirements

› [More](#)

is dramatically reduced from what it was in the initial release of Exchange 2007. Formerly, database I/O could be as much as 200% to 300% more on replication targets than it was on the source. In SP1, database I/O on the target is now 78% less than on the source. Target log I/O has also been reduced in SP1.

**The aggressiveness of online defragmentation has been reduced and monitoring is provided.** You can now see how long a defragmentation pass takes to complete and how often the database has been fully defragmented. You can use this information to adjust the online maintenance window to reduce disk churn, which affects Snapshot™ space consumption.

**Database scanning can now be performed during online maintenance to check for corruption.** Prior to SP1, the only way to ensure that a database was not corrupt without taking it offline was to perform an online streaming backup, forcing each page in the database to be read and verified. This was also the only way to force the zeroing of deleted pages, as mandated in some security scenarios. If you've been performing VSS backups on a replica rather than on the active database, this probably means that your production database has never been fully verified.

In practice, both NetApp and Microsoft recommend using Visual SourceSafe® (VSS) for backup. The backup application verifies the backup data, but once again a check is never actually run on the live data. That creates a small but real risk of database corruption creeping in.

When enabled in the registry of an Exchange server with SP1, database scanning and page zeroing run every day during the online maintenance window to verify the active database and zero deleted pages.

### Best Practices

At this point you're probably saying, "That's all great, but how should I take advantage of these new features in my NetApp environment?"

**Windows Server 2008.** If you want your Exchange servers to run the latest OS, you need to install SP1 on a new server running Windows 2008; upgrading the OS of a running server is not supported. You don't need to install Exchange 2007 before installing SP1; the service pack includes everything.

Regardless of OS, I recommend upgrading to SP1 for the bug fixes and enhancements, which have a major impact on storage.

**LUN configuration for replication.** The Microsoft recommended best practice is that you provide the same LUN configuration for both the source and target sides of each replica—and naturally you'll want the source and target LUNs to be on separate storage systems. Because of the heavy I/O load that the initial release of Exchange 2007 put on target servers, we used to recommend that you configure a separate log and database aggregate for each cluster node. If a storage system was supporting multiple replication servers (sources and/or targets), it needed a lot of independent aggregates.

With the reduction in I/O load in SP1, that requirement goes away. We continue to recommend that you keep logs and databases in separate aggregates (and you should still keep source and target storage separate), but you can now group multiple databases from separate clusters into a single aggregate, and the same goes

### Storage Strategies for Exchange Server: Providing a Highly Efficient Enterprise Messaging Infrastructure

In this On Demand Webcast, industry experts discuss tools and strategies to reduce infrastructure costs and achieve a highly available enterprise messaging environment with Microsoft Exchange Server 2003 or 2007.

Learn how to:

- Develop a solid transition plan for pain-free Exchange Server migration and consolidation.
- Maintain availability of your data during planned or unplanned downtime.
- Reduce storage costs by more than 36%.

[More](#)

for logs. This eliminates “islands” of storage, simplifies storage configuration, and can increase utilization as well.

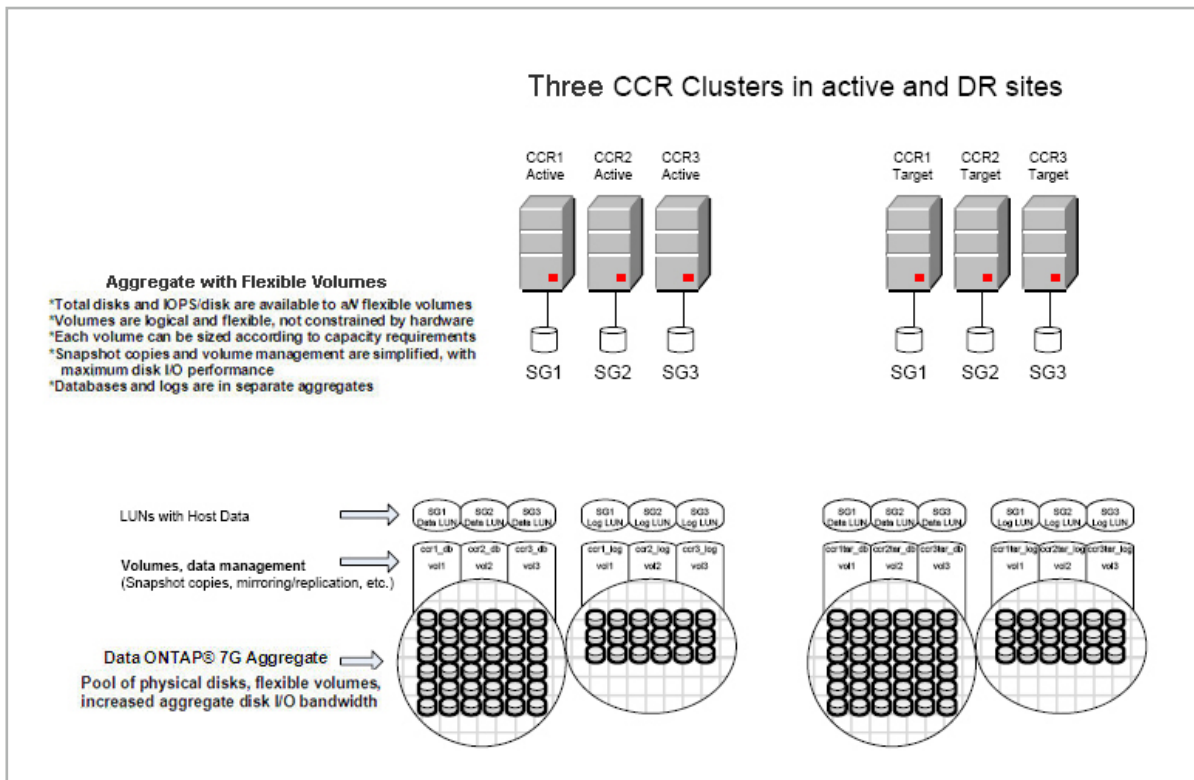


Figure 1) LUN configuration for replicas with Exchange 2007 SP1. Both source and target storage systems use 1 aggregate for databases and 1 for logs for a total of 4 aggregates (2 on each system) versus the 12 aggregates that were previously needed.

**Online defragmentation.** Defragmentation can have a big impact on the number of blocks that must be retained by NetApp Snapshot copies. The defrag process changes blocks, and NetApp Snapshot copies maintain a stable, point-in-time image by retaining change blocks. Therefore the more blocks you change through defragmentation, the more space you need to store Snapshot copies.

This is where the new metrics come in. Microsoft recommends that defragmentation be completed in 14 days, so if you check the metrics and find that defragmentation is completing much sooner, you can shorten the length of the daily online maintenance window to reduce the amount of daily disk churn, which in turn reduces Snapshot space consumption.

For more precision, two performance counters can be logged to determine your online defragmentation trend:

- MExchange Database ==> Instances\Online Defrag Pages Freed/Sec
- MExchange Database ==> Instances\Online Defrag Pages Read/sec.

If the ratio of reads to freed pages is greater than 100 to 1, reduce the online maintenance window.

If the read-to-freed ratio is less than 50 to 1, increase the online maintenance window.

**Database scanning and page zeroing.** By default, database scanning is either on or off. If it's on, it runs every night during your online maintenance window. That's okay, but you have to ask yourself whether you really need to verify your database that frequently, since essentially it reads the entire database block by block. Naturally, that creates a lot of extra load on storage systems.

If page zeroing isn't critical to you and you have NetApp SnapManager® for Exchange, an alternative is to configure a weekly copy backup that doesn't truncate your logs. This will check for corruption and verify your

active database. You may be able to schedule it on a weekend when the workload is lighter.

If you absolutely have to have page zeroing, running database scanning every night is the best way to ensure that it happens. Be aware that if a database didn't have page zeroing enabled from the beginning, the first time it runs it will create an extreme I/O load. To reduce the impact of the first pass of page zeroing, you can enable throttling.

### **Getting the Most from Exchange 2007 SP1**

Exchange 2007 SP1 offers so many enhancements and new functionality that it's almost like a new release of Exchange. By paying attention to a few best practices, you can get the most from your Exchange installation and your NetApp storage. Here's a summary of current NetApp best practice recommendations for Exchange 2007 SP1.

### **LUN Configuration and Replication**

- Isolate active and target server storage on separate storage systems.
- Provision the active and target LUNs identically with regard to capacity and performance.
- Separate logs and databases in their own aggregates.
- Create a separate NetApp FlexVol® volume for each storage group.
- Use NetApp RAID-DP® for superior performance and protection.
- Run SnapManager for Exchange backups on the target node and adjust the online maintenance window on the active node.
- Consider NetApp ReplicatorX™ and/or SnapMirror® to achieve an RPO of less than 5 minutes when replicating Exchange databases, logs, and hub transport data.

### **Defragmentation and DB Scanning**

- Reduce the online maintenance window if it is completing a full pass within 2 weeks and the read-to-freed ratio is greater than 100 to 1.
- Consider using a copy backup (which does not truncate logs) and checksum integrity to validate database health versus online database scanning.



#### **Robert Quimbey**

Microsoft Alliance Engineer  
NetApp

Robert joined NetApp in 2007 after 8 years as a member of the Microsoft Exchange product team, where he was responsible for storage and high availability. Since joining NetApp, his activities have continued to center around Exchange 2007, including designing best practices and an in-depth analysis of Clustered Continuous Replication (CCR) I/O. His recent work will lead to the creation of reference architectures for Exchange.

---

**Interested in learning more about deploying Exchange on NetApp?** Visit NetApp at Microsoft TechEd in Orlando, Florida, June 10 - 13, 2008. Stop by booth 201, and attend our technical session.

#### **Wednesday, June 11**

2:45-4:30 p.m.

How resilient is your Microsoft® applications infrastructure? Corporate IT organizations struggle with business continuity and disaster recovery challenges every day. Often there is no simple solution, and the current strategies are usually untested. This session discusses how you can address business continuity and disaster

recovery today with Microsoft products and technologies.

At this session, you will learn how to create high-availability and disaster recovery solutions for Microsoft applications such as Exchange Server, SQL Server™, Office SharePoint® Server, and Microsoft Hyper-V™ to ensure that your next disaster recovery exercise will be a positive and defining moment in your career.

---



# Tech OnTap

## Quick Links

- › [netapp.com](http://netapp.com)
- › [Tech OnTap Archive](#)

## Simplified Upgrade Planning with Premium AutoSupport

By [Jay Mistry, Alliance & Leicester](#)

One of the biggest hassles for storage administrators is keeping on top of needed upgrades and general maintenance. My company, Alliance & Leicester—a UK financial services company—has been a NetApp customer for about seven years. During that time, we've grown our midrange storage pool for our Windows® environment to about 25 storage systems with 130TB of stored data and 150TB of total capacity, so over time management complexity has grown.

NetApp storage has worked well for us, and we've been extremely happy with overall system reliability. Ease of management makes it possible for us to take care of our entire NetApp® environment with just one full-time person (me) plus two additional people who assist me part time. Still, I used to spend quite a bit of time doing my upgrade planning, and I have always had a hard time keeping track of e-mail alerts that advised me of needed changes to my systems.

Premium AutoSupport from NetApp Global Services addresses these and a variety of other challenges by simplifying and automating what were previously manual processes. If you haven't had a chance to see it, I would highly recommend that you [check out the demo](#). Tech OnTap asked me to describe our experiences with Premium AutoSupport for this article.

### NetApp Premium AutoSupport

When you buy a Premium support contract on a NetApp storage system, you get Premium AutoSupport as part of the service. Every NetApp system is configured with AutoSupport, a call-home service that provides system alerts and sends weekly logs to NetApp. Premium AutoSupport extends AutoSupport to create a proactive self-service support tool that gives you a lot of useful functionality, including the ability to:

## Explore

### About Alliance & Leicester

Alliance & Leicester is one of the UK's major financial services groups, offering a broad range of services to personal and commercial customers. It has been a member of the FTSE 100 index of leading shares since 1997.

› [More](#)

### Increase Operating Efficiency

Studies show that up to 60% of problems requiring support are due to configuration issues. A recent NetApp TechBrief featuring this month's author, Jay Mistry, explains how Premium AutoSupport can:

- Reduce system support overhead
- Improve uptime
- Increase operational efficiency

[This on-demand Webcast](#) includes a detailed discussion of the product along with a full demonstration.

### Five Tips to Increase NetApp Resiliency

A recent Tech OnTap article examined five little-known things that you can do to improve the resiliency of your existing NetApp storage systems, including:

- Multipath HA
- Proper sparing
- SyncMirror®
- Nondisruptive upgrade

- Visualize NetApp systems at the company or device level.
- View a timeline of AutoSupport events and alerts.
- Compare the differences between two configurations.
- Generate an upgrade plan from one Data ONTAP® release to another.
- View proactive recommendations for your system (health check).
- View raw AutoSupport data content.
- Customize AutoSupport e-mail notifications.

- Automated tools

[More](#)

These features are all accessed from a Premium AutoSupport dashboard hosted on the [NOW™ \(NetApp on the Web\) Web site](#).

At Alliance & Leicester we've been using Premium AutoSupport since it first went into beta test in September 2007. We've come to rely on three features of Premium AutoSupport: 1) upgrade planning, 2) event and alert history, and 3) system health checks.

We currently use NetApp Operations Manager for day-to-day management of our NetApp storage. The real-time monitoring of Operations Manager complements Premium AutoSupport's ability to handle surface configuration and upgrade issues. If we weren't using Operations Manager and other tools in house, we'd probably use several more of the features of Premium AutoSupport. If you don't have tools for visualizing your storage configurations or comparing systems, you should definitely check out those features too.

### Keeping Up with General Maintenance

The ability to view noncritical events and alerts in one place when I have the time to address them works much better for me than trying to deal with AutoSupport e-mails that arrive in my mailbox at random times.

Once a month or so I log into the NOW Web site, go to Premium AutoSupport, and look at the timeline of alerts on each of my storage systems. I can easily drill down and see specific details for each item (Open Cases, Hardware Failures, System Alerts).

I also use the system health check feature to analyze my system configurations and identify configuration issues—such as out-of-date firmware or overused snap reserves—that may affect stability. With all this information at my fingertips, I can put a plan in place to take any actions that are necessary. (You can view a sample of health check output below.)



Figure 1) Event history.



Figure 2) Health check.

### Planning Upgrades

I used to spend a lot of time on the NOW Web site manually checking firmware revisions, patches, and so on and looking through upgrade planning documents in order to create a plan for each of my storage systems. While I never experienced any problems as a result of this method, it's obviously time consuming and there is some opportunity for error.

Now, Premium AutoSupport's Upgrade Advisor tool does all the hard work for me, saving me four to five hours of work per storage system and providing an upgrade plan that's complete and easy to follow. The tool works for both standalone systems and clusters and supports all Data ONTAP releases since 6.3. Among the things that it checks for are:

- Known upgrade issues
- Required firmware revisions
- Applicable field alerts
- Software license issues
- Version compatibility of dependent software products (SnapDrive®, etc.)

Another advantage of the Upgrade Advisor is that it gives us a complete plan to back out each upgrade should a problem arise. We've never had to use such a plan, but new regulations in the finance industry are starting to mandate this capability, so it's nice to have it covered.



Figure 3) Upgrade advisor.

### What's Next?

Over time, the applications we support—Oracle® and FileNet on UNIX®, Microsoft® Exchange, plus a variety of applications running on Microsoft SQL Server™ and Oracle for Windows—are becoming more and more mission critical. Our internal change management policy has some very specific mandates, so we currently have to start planning several weeks in advance in order to schedule any upgrades.

Going forward, we'll be moving to larger, clustered NetApp storage systems all with Premium support. This will give a higher guarantee of availability for our critical applications and will also make nondisruptive upgrade (NDU) possible.

With NDU, we'll be able to upgrade one member of a cluster at a time without disrupting running applications or user access. This will reduce the upfront scheduling we have to do to accomplish an upgrade while increasing our service level. I understand the next release of Upgrade Advisor will include NDU checking and recommendations, so upgrade planning will remain straightforward as we move ahead.



### Jay Mistry

Senior Telecommunications Analyst  
Alliance & Leicester

With 20 years of IT experience, Jay has been employed at Alliance & Leicester for the last 11 years supporting the company's Windows server environment. He became involved with storage, initially with HP SANs, more than eight years ago and has been working with NetApp storage for four years. Today Jay is responsible for 25 NetApp systems with about 150TB of total storage.



# Tech OnTap

## Quick Links

- [netapp.com](http://netapp.com)
- [Tech OnTap Archive](#)

## 100 Year Archiving Are You Ready?

By [Gary Zasman](#)

Pick up a book from 100 years ago and you can probably read it without much trouble, but pick up a backup tape that's just 10 or 20 years old and it's likely to be much more problematic. Even if you have the right hardware to read the tape (and the tape itself is still good), you need to know what format it was written in, and you need an application that can make sense of the data. The situation can get very complicated, and it only gets worse as time passes.

At this point you may be thinking, "That's okay, I don't need to keep data that long." Think again.

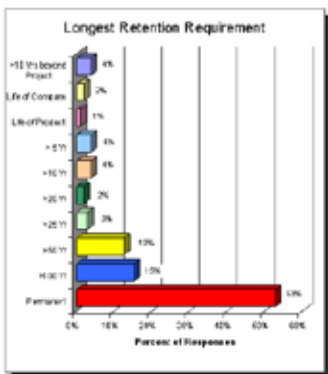


Figure 1) Data retention requirements.

Source: [SNIA Survey](#)

In 2007, the Storage Networking Industry Association (SNIA) completed a [comprehensive survey](#) in which we talked to hundreds of individuals in a wide variety of organizations from countries around the world. An astonishing 80% said they have information that they must keep over 50 years, and 68% said they must keep data over 100 years. A full 70% also reported that they are highly dissatisfied with their perceived ability to read retained information in 50

years. What type of data were they most concerned about? E-mail, customer records, business application data, and databases—the kinds of information that most of us deal with every day.

Have I got your attention?

In this article, I'll explain the challenges of long-term archiving, discuss a few best practices that you can use now, and talk about efforts that are under way through the

## Explore

### Migration, Emulation, or Encapsulation?

There are currently three methods of protecting data when the hardware or software that you use needs to be managed for obsolescence: migration, emulation, and encapsulation.

**Migration** simply means copying digital files between storage subsystems. Migration is the most cost-effective and widely practiced method of preserving digital data. This approach may sound simple, but it does have disadvantages.

- Repeated migration runs the risk of losing or corrupting data or being locked into proprietary systems.
- Because digital technology is developing almost at the speed of light, it is impossible to predict how frequently migration will be necessary.
- Because of the proprietary nature of some commercial software and hardware, migration of digital data may raise interoperability and chain of custody issues.

**Emulation** involves creating software that mimics, or emulates, the behavior of an outdated computer. This approach avoids the need for continuing conversion of digital records, but emulators must be created for every software/hardware configuration, which makes it far more expensive than migration. It is favored by some experts in the preservation field,

SNIA Long Term Archive and Compliance Storage Initiative (LTACSI), which I chair.

### What Are the Long Term Archiving Challenges?

Figure 2 makes clear the challenge of long-term data retention.

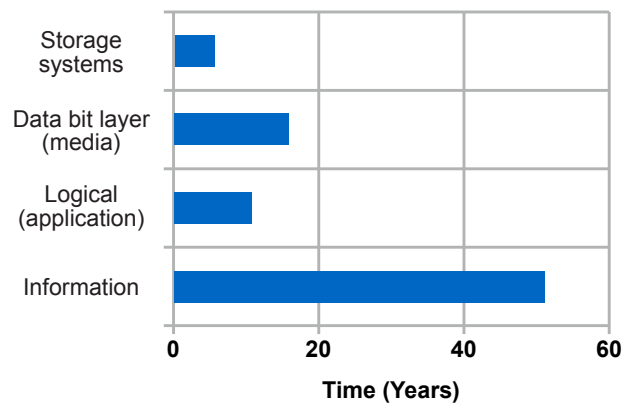


Figure 2) Typical lifetimes of storage systems, applications, and physical media versus information retention.

Simply put, the time that we need to retain information (even using the more modest figure of 50 years) far exceeds the typical lifespan of storage systems (disk or tape) and applications. Even the physical media start to degrade and may become unreadable long before the retention period expires.

The current practice is to migrate data—both physically and logically—every 3 to 5 years. Physical migration requires moving information from one physical storage system to another or from one media format to another to maintain physical readability, accessibility, and integrity. Drivers for this type of migration include media failure, media or storage system obsolescence, system changes, and cost of operations (people, power, space).

Logical migration requires moving information from one logical format to another—such as from an old version of an application to a new version—to preserve readability and interpretability. Drivers may include changing application formats, obsolete applications, and mergers. Inhibitors to both types of migrations include cost, complexity, sheer volume of information, and lack of time and/or budget.

The SNIA survey mentioned previously concluded that logical and physical migration simply do not scale cost effectively to meet current and future needs. In fact, only 30% of those surveyed migrate data every 3 to 5 years if it's on disk. Clearly, new approaches are required to meet the legal, regulatory, business, cost, and scalability requirements for long-term digital information retention.

and some emulators like VMware can be used.

**Encapsulation** groups a digital object with all other associated entities that are necessary to provide access to that object. In encapsulation, physical or logical structures called "containers" or "wrappers" provide information about the relationships between all data and software application components.

Encapsulation aims to overcome the issue of obsolete file formats by including details on how to interpret the original information. The SNIA LTACSI's proposed encapsulation format—modeled on the OAIS AIP—has been initiated to explore encapsulation for corporate archives, but it is still a developing strategy and will not be practical for some time.

Until encapsulation technology matures, NetApp recommends using open, standards-based storage that supports open interfaces and virtualization. This will give you many more options in the future—whether you choose migration, emulation, or encapsulation—and also protects you against obsolescence and vendor lock-in.

### Advanced Technology at NetApp

When it comes to future technologies, NetApp maintains an aggressive, multipronged portfolio of investments.

In addition to its efforts to drive the standards that will shape the future of storage, NetApp has an Advanced Technology Group (ATG) that is looking out 18 months to as much as 5 years—beyond the next several product releases—to do speculative technology development and research that may be useful in the future.

Support for AT projects in NetApp Product Development is another aspect of the portfolio. ATG acts as an internal

## Interim Solutions

The limitations imposed by current storage systems and applications are not going to go away in the near term. What then should you do today to address long-term retention? The best current recommendation is to implement formal lifecycle management processes for your applications, operations, and data repositories to address the effective management of data through its useful life. Best practices should include:

- Close collaboration among all stakeholders (IT, RIM, legal, business, security) to ensure that all needs are addressed
- Clear identification of all existing assets and resources
- Classification of information so that retention needs can be determined
- Establishment of requirements for retention, protection, security, compliance, and so on
- Service implementation to meet requirements
- Measurement and improvement

Helpful practices may include:

- Classifying information into a few common buckets
- Setting retention periods and deleting expired data
- Controlling the number of copies of data that you maintain for data protection
- Setting policies for audits and performing them
- Using standards-based storage platforms

Your long-term preservation policy should identify your business, legal, and compliance goals and include a description of the best practices to which each storage repository adheres, including both physical and logical migration. The goal for physical migration should be to move from fixed-term (3 to 5 years) migration to an “as needed” strategy. Federated, standards-based, and virtualized systems (such as NetApp® storage running Data ONTAP® GX) can help minimize the disruption, complexity, and labor involved in migration.

For logical migration, you must be able to maintain authenticity—proof that the data is what it was originally. Again, you should migrate only as needed, and you may want to consider other options for retaining some data: transform it into a standard format (XML, PDF, etc.), archive a hard copy if appropriate, or use microfilm.

If this seems complicated, it is. You still have to execute both physical and logical migration on an as-needed basis, and the two events may not coincide. But there are currently few options to ensure that data retained over the long term will remain readable. Fortunately, some significant efforts are underway to help address this situation.

venture capital group, soliciting proposals from the product development community and funding those that have business value.

The third aspect of AT investment is university research. NetApp funds research at leading universities that are doing research in storage and systems, including University of Wisconsin, the University of California at Santa Cruz, Harvard, Carnegie Mellon, and others.

Looking at the complete portfolio, ATG is a riskier investment with a longer-term return. Advanced technology work in engineering gives us more near-term returns and not as much risk; university research is more speculative.

### Collaborative Research into Long-Term Archive Systems

One joint research project between NetApp and the University of California at Santa Cruz focuses on the creation of low-power storage systems that are capable of lasting for decades.

The project draws on commodity components to create hardware modules called “tomes” that are joined together in a grid with self-healing properties.

[➤ More](#)

### Long-Term Archive and Compliance Storage Initiative (LTACSI)

LTACSI is a cooperative effort of end users, IT professionals, vendors, integrators, and service providers with interest in the challenges of long-term digital information retention, archiving, and compliance-related storage practices.

[➤ More](#)

## Standards Efforts

It's probably clear that the storage industry hasn't done a lot to address the problem of long-term data retention thus far. Today's archiving applications use proprietary data formats that effectively lock you into a solution and that could further complicate migration efforts in the future. That's all about to change.

For long-term archiving (15+ years), the biggest challenge is logical migration. The physical migration situation can be adequately addressed with effective lifecycle management processes and current standards-based storage technology, as opposed to proprietary storage formats. The situation should improve further as vendors begin to focus more attention on hardware that can meet long-term storage needs. (For an example see the sidebar, *Collaborative Research into Long Term Archive*.)

Logical migration, on the other hand, remains application-specific, making automation of key processes more difficult. Full "preservation" requires more than just keeping the data readable and interpretable, it requires long-term retention of the data with metadata that includes its provenance, its reference information (context), and mechanisms to ensure its integrity and authenticity.

To this end, the SNIA LTACSI proposed that SNIA form a Long-Term Digital Information Retention and Preservation Technical Working Group to look into encapsulation (see sidebar). Encapsulation would define a "preservation-oriented" logical container consisting of the content (the data) and associated preservation metadata.

Encapsulation could be modeled on the OAIS AIP (Archival Information Package). Figure 2 shows the content of an OAIS AIP container.

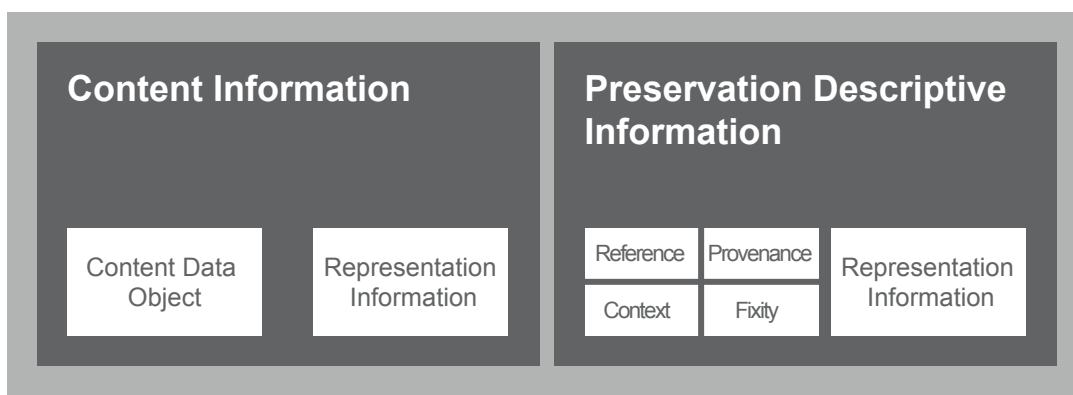


Figure 3) OAIS AIP includes both the content of the information to be stored and metadata describing that content. (Source: SNIA)

Encapsulation implies "self-contained" because a container holds the information's data, its metadata, reference information, integrity and authenticity checks, access controls, and logs. This content makes the container portable and independent of storage. It allows the container to be managed independently of the application, in accordance with the requirements you've established for the information.

Encapsulation is "self-describing" because the container could be interpreted by different types of systems and because it can include readers, allowing the contents to be interpreted independently of the application. This capability is important for long-term preservation. Encapsulation provides a standard format that any application can understand and that in theory allows many application types to access archived content, such as ECM, legal, migration, preservation, and so on.

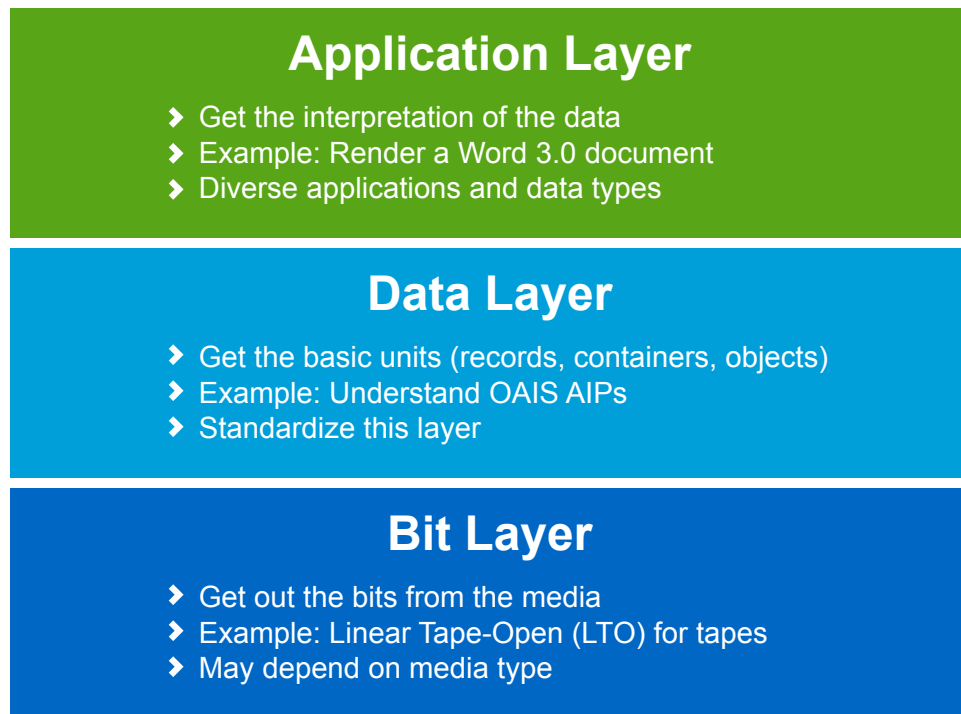


Figure 4) Logically, encapsulation creates a standard data layer that fits between the bit layer (physical media) and the application.  
(Source: SNIA)

The goal is to eliminate the need for frequent logical migrations so that organizations can continue to access and use archived data as necessary over long periods without the overhead and complexity of regularly updating data to accommodate application changes.

### NetApp and Long-Term Archiving

From a hardware perspective, NetApp has long recognized that physical data migrations—whether they are for archival or other purposes—are complicated and disruptive. For this reason, NetApp is moving toward a scale-out hardware architecture that supports tiered storage—including write once, read many (WORM) volumes for compliance needs—to ease migration of data from one tier to another in a nondisruptive manner. This architecture allows the transparent incorporation of new storage building blocks (physical media, storage controllers) alongside existing storage, greatly simplifying the physical migration process.

To meet archive and compliance requirements, NetApp open SnapLock® technology enables the creation of WORM volumes on NetApp storage to meet corporate governance and regulatory requirements without requiring physically separate storage systems. NetApp works with industry-leading archive partners, such as Symantec, Zantaz, and CommVault, to deliver solutions that leverage the unique features of NetApp hardware and software technology and is also collaborating with these partners on long-term solutions.

From the standpoint of logical migration, NetApp knows that solving near-term archiving problems is only part of the solution. We recognized the need for industry standards early and have been a key contributor to standardization efforts. In my role at NetApp, one of my chief responsibilities is to chair LTACSI, a cooperative effort of end users, IT professionals, vendors, integrators, and service providers with interest in addressing the challenges of long-term digital information retention, archiving, and compliance-related storage practices.

### What Should You Be Doing Now?

The most important thing is to begin taking steps now to avoid ending up in a crisis situation with terabytes of data requiring physical and/or logical migration. The best way to do this is to follow the guidelines described in “Interim Solutions,” including the use of open standards wherever possible. Open standards give you many more options when it comes to migration and help prevent lock-in.

If your organization has not done so already, consider implementing data classification to better understand your data and to support lifecycle management. Then look for solutions, both hardware and software, that can enforce policies and simplify the physical migration process. By taking these steps now, you'll be well positioned to take advantage of new long-term archival standards as they take shape in coming years.

---



**Gary Zasman**

Worldwide Practice Director  
NetApp

Gary serves as the chair of the SNIA Long Term Archive and Compliance Storage Initiative (LTACSI). He also spearheads the development of the NetApp worldwide practice for business applications and database integration. Before joining NetApp in 2006, Gary held a variety of positions with leading storage vendors focused on the development of ILM solutions and consulting practices. In 2001, a team that Gary worked with was selected as a finalist for the prestigious Computerworld Smithsonian Award for developing a visual history digital archive.

---

**Interested in learning more?**

Visit [NetApp at Symantec VISION](#) in Las Vegas, Nevada, on June 10 through 12. Stop by NetApp booth #201, and be sure to attend our technical sessions:

**Birds of a Feather Session**

Tuesday, June 10, 1:00 p.m.

Join leaders from ING, Countrywide, and Renault F1 to learn how they implemented EV, NBU, and VCS with NetApp to go further, faster.

**Technical Sessions Featuring NetApp Customers**

Wednesday, June 11, 9:00–10:00 a.m.

Can You Find That E-Mail? Legal E-Discovery for Hospitals

Thursday, June 12, 9:00–10:00 a.m.

Disk to Disk to Tape for the Small Business Customer

Thursday, June 12, 10:15 a.m. – 11:15 a.m.

NetBackup OpenStorage API: Cutting Edge Support for Backup Appliances

Thursday, June 12, 11:30 a.m.–12:30 p.m.

Reliable Enterprise Application Hosting with Symantec NetBackup and NetApp VTL at T-Systems

---