SAP’s Strategy for End-to-End ILM Success

The Information Lifecycle Management Solution from SAP Bridges the Gap Between Applications and Storage Technology for Legal Compliance

Ensuring that your company’s data management meets internal and external regulatory, governance, and business requirements has become an issue that, if mishandled, can cost millions of dollars.\(^1\) As companies confront this challenge, they may also find that their current policies and traditional approach to data management prove insufficient to handle the complex demands of new compliance requirements.

Information lifecycle management (ILM) is a response to just this — and is an evolution of traditional data management. If you remember from a previous Performance & Data Management Corner column,\(^2\) ILM is about more than just managing data; it’s about handling information, which includes the related business context and restraints that are so crucial to successful governance.

This means that managing information must start well before your data goes into archive storage. SAP strongly advocates that ILM can only be effective with a holistic approach that covers data from the moment it is created in SAP to the moment it is destroyed — often in a third-party storage system.

Successful ILM should span applications, storage systems, and even interfaces and protocols; installing the right tools and setting policies for data retention and destruction is just a start. You must ensure your business systems can communicate the right information — not just the data — to your storage systems. With SAP’s approach to ILM, this is easy.

The earlier Performance & Data Management Corner article detailed the basic ideas behind effective ILM and SAP’s approach, and it introduced a new tool from SAP due out in 2008: the Information Retention Manager. This tool will become the central policy engine for creating and managing retention rules for all types of information. But once you’ve set your retention requirements within SAP, how will this information be applied to the appropriate data in your storage system so that information is correctly archived, retained, and destroyed there?

This article addresses this question and introduces a key equation for SAP customers looking ahead to successful ILM:

Information Lifecycle Management from SAP = SAP applications + ILM-aware storage

What Is ILM-Aware Storage?

In SAP’s ILM strategy, a storage system must “understand” and apply retention rules originating in the SAP application. More precisely, the storage system must accept the results of the retention rules (in particular, an expiration date calculated by the SAP system based on a beginning date and retention period) and enforce them on the archived data. SAP is working closely with different storage partners to achieve just that.

Storing data, especially large files — in read-only mode at low cost and over a long period of time — has always been one of the strengths of SAP storage partners. However, recent innovations in storage technology have emerged, which set the stage for achieving SAP’s ILM approach. These innovations include storing data immutably even when using magnetic disks and commodity hard drives. Examples include the concept of Content Addressed Storage (CAS) in EMC’s Centera — and similar products from NetApp, HP, or IBM — with WORM-like storage behavior, even with read/write media.

These newest types of storage products also ensure that data cannot be destroyed before its creation. However, managing your company’s information in such storage systems is not straightforward.

\(^1\) See Network World articles “Sloppy e-discovery can cost you millions” by Joanne Cummings (May 21, 2007) and “E-discovery law a boon for lawyers” by Ellen Messmer (October 19, 2007).

\(^2\) See “From Data Management to Information Lifecycle Management” by Dr. Bernhard Brinkmöller and Georg Fischer in the July-September 2007 issue of SAP Insider (www.SAPinsideronline.com).
expiration has been reached or before all legal holds have been lifted — provided that these retention constraints are communicated appropriately. And this holds true no matter where data is being accessed from, whether it’s from an SAP system or any other (potentially untrustworthy) client application.

Why ILM Starts in Your Business Applications

A major part of ILM is making sure that you abide by data retention laws and regulations. One of the most dramatic rulings is the much-heralded US federal “e-discovery law” that went into effect at the end of 2006.1 This law, along with other compliance mandates, is motivating companies to ensure their electronic data is transparent, auditable, easily manageable, and accessible (see sidebar).

The trick is that defining retention rules and managing legal holds — key processes of a company’s data compliance strategy — involve operations that are primarily seen from a business point of view. The rules are formulated independently of a technological implementation, regardless of whether the data is kept in the database or a third-party storage system. Accordingly, your SAP system plays a natural leading role in controlling the life cycle of business information. After all, the SAP system is where most business knowledge is modeled. Most SAP business objects start their lives in the database of the SAP system and, over time, are transferred to long-term storage through data archiving before they are finally destroyed.

Consider legal retention time, for example. For compliance, the storage system must use the retention information it receives from the application — any time stamp the storage system assigns to the object at the time of archiving is completely irrelevant to the application-defined retention period.

So the first issue is this: How do you ensure that the storage system can apply the retention rules it receives from the business side of the equation? And then there’s the technical integration issue: With so many proprietary interfaces for well-established storage systems,2 how do you bridge SAP applications and third-party storage systems to achieve comprehensive ILM?

How SAP and Storage Partners Are Tackling ILM Challenges

On the SAP side, these issues are addressed with open interfaces in SAP NetWeaver 7.0 and SAP ERP 6.0, which make the most of ILM-aware storage without restricting integration to any proprietary API of a storage system. The goal is to provide SAP customers with as much flexibility in choosing their storage systems,2 while proprietary interfaces are the norm, there is a standardization initiative in draft status within the nonprofit Storage Networking Industry Association (SNIA) to harmonize storage APIs; it’s called the eXtensible Access Method (XAM). It focuses on storage layer objects, basically BLOBs in a flat address space — without hierarchical structuring — and attributes. SAP envisions XAM could become a technical interface “below” the open ILM interface described in this article, helping third-party vendors in particular to interoperate with different storage back ends.

New E-Discovery Law Requirements: Legal Holds, Data Retention, and Immutability

The 2006 US federal e-discovery law requires companies to quickly put any data destruction on hold (known as a “legal hold”) as soon as a suit is filed against them. To be legally compliant, you must be able to prove that information was not tampered with and ensure that important information is not inadvertently deleted before its time. Auditors and attorneys check for technology and system capabilities that guarantee this for the data, both in the application and the storage system. In these cases, taking policies set up for IT-managed business objects and translating them to manage the life cycle of the right storage objects is the big challenge.

And data retention is not limited to US law: Depending on the country, companies are required to keep sales orders, for example, for different lengths of time. What’s more, it makes no difference whether the sales orders are scattered across database tables or stored as binary large objects (BLOBs) in a storage system, how redundantly they are stored, or in how many formats.

The retention rules for these orders can be quite complex, involving far more than simply calculating the expiration date based on a single criterion. Rather, the calculation involves many criteria, such as document category, company code, and sales organization, and depends on different types of starting points for retention, such as change date or end of fiscal year. In addition, some objects inherit their retention policies from other superordinate objects. These are just some of the challenges that the ILM solution from SAP is designed to address.

---

1 The Federal Rules of Civil Procedure were changed to ensure that companies can quickly and transparently hand over electronic communications related to federal cases — or put a hold on any purging or destruction of e-communications — as soon as a suit is filed, often well before the case goes to trial. For more information, see www.networkworld.com.

2 While propriety interfaces are the norm, there is a standardization initiative in draft status within the nonprofit Storage Networking Industry Association (SNIA) to harmonize storage APIs; it’s called the eXtensible Access Method (XAM). It focuses on storage layer objects, basically BLOBs in a flat address space — without hierarchical structuring — and attributes. SAP envisions XAM could become a technical interface “below” the open ILM interface described in this article, helping third-party vendors in particular to interoperate with different storage back ends.
vendor as possible. SAP has put together three elements to bridge the gap between your SAP application and an ILM-ready, third-party storage solution, some of which will be familiar to SAP customers (see Figure 1):

- **Information Retention Manager (IRM)**, for managing retention policies and providing the resolved retention result at runtime
- **Archive Development Kit (ADK)**, for the mediator that binds the IRM output to ADK files and determines placement in the archive hierarchy (see sidebar on page 62)
- **XML DAS with ILM-enhanced WebDAV interface**, for moving the data, along with its associated retention information, to the storage system

### Mapping Retention Rules onto a Storage Hierarchy

IRM is the central policy engine in which users can define retention-relevant attributes of business object types (e.g., fields of Business Object Repository types or archiving objects). They can then use these attributes to assemble retention rules, which set the residence and retention times for object instances.

But there is more to a retention rule than when and how long data should be stored in an archive. At runtime, the SAP system takes the values of the retention-relevant attributes and concatenates them into a URL-like path (see Figure 2) so that a hierarchy is built based on the retention criteria you entered in IRM. Figure 2 shows a node used for all financial documents (FINDOCS) from a system called SID, with company code USA, vendor 7, and fiscal year close in 999. The archive hierarchy’s path is thus /SID/FINDOCS/USA/7/999. This type of structure has two key benefits:

- **The hierarchy provides an easily traceable organization of stored data for many years to come.** Browsing this hierarchy, you could locate business object instances with the same retention rule under the same nodes. For example, with any lawsuit that involved financial documents of vendor 4711 with the criteria in Figure 2 for fiscal year 1999, you could place a legal hold on this single node.

---

5 The definition of the “start of retention period” leads to the creation of the corresponding node “1999” in our example.
node. Likewise, when the expiration date has been reached, this single node can be deleted to destroy all archived objects it contains.

In addition, if you are not interested in the retention aspect, the structure of the hierarchy still presents you with an intuitive way to locate information, even if the original system has been shut down and corresponding indexes no longer exist.

- This structure allows IRM rules for an object to be inherited by its subordinate objects and even be applied to corresponding attachments. To continue our example, the legal hold on the 1999 node would still apply, regardless of whether any child collections exist that represent different accounts below 1999. In other words, the legal hold is inherited by all archived financial documents below the node 1999.

### Using WebDAV as a Platform-Independent, Standard Protocol

Open standardized protocols that support hierarchies are hard to come by, especially when you also want to pass metadata (attributes or properties) along with the data you are storing. Therefore, SAP decided to evolve its own WebDAV-based archiving storage interface (first available as part of the XML Data Archiving Service, or XML DAS, in SAP NetWeaver 2004) into an ILM interface between the application system and the ILM-aware storage system.

Based on the WebDAV standard concept of “dead” and “live” properties, SAP is able to predefine properties to pass an expiration date to, or place a legal hold on, those storage objects that represent the business objects.

In the case of the expiration date (see Figure 3), the SAP system calculates it using the applicable retention rule based on the policies set in IRM. Any standard WebDAV server should accept these properties, but only the ILM-certified storage vendors assert that their solutions will understand them. ILM-certified storage systems turn the assigned expiration date into guaranteed nondeletability of the storage object as long as this date has not yet passed. In other words, the dead property becomes a live one for ILM-certified WebDAV vendors, and the retention constraint is ultimately enforced.

The mechanism for setting legal holds works similarly (see Figure 4 on the next page). The ILM-aware storage system accepts multiple legal hold cases for a given storage object in the form of an XML-valued property of a predefined namespace and property name. Because of the retention policy-driven hierarchy, an expiration date assigned to a node (called collection in WebDAV) affects all child objects and the contained business object instances (called resources). Usually, you can propagate legal holds hierarchically as well, so that one WebDAV request issued by the SAP system is sufficient to freeze all related information, such as attachments of a business object instance. Since the ILM-aware storage

### Is Your Storage Partner Ready for SAP’s ILM Approach?

First, be sure to ask your storage partners about their current SAP certification. The most recent certification for the WebDAV interface through the SAP integration scenario is “BC-DAR 1.1,” which includes the WebDAV standard’s property handling.

Then, ask about their future plans for SAP certification with the ILM solution from SAP. The 1.1 certification paves the way for full-fledged, ILM-aware certification to be available in the first half of 2008.

For more information on WebDAV certification, see [https://www.sdn.sap.com/irj/sdn/ic#Integration Scenarios](https://www.sdn.sap.com/irj/sdn/ic#Integration Scenarios).

---

1 WebDAV, which stands for Web Distributed Authoring and Versioning, is published by the Internet Engineering Taskforce (IETF) and is an extension to the HTTP protocol. WebDAV is particularly valuable because of its standardized creation of hierarchies in the form of nodes for content resources and for its storing and retrieving of attached properties. For more information, visit [www.webdav.org](http://www.webdav.org).

2 “Dead” means the server merely records the value of the property without any processing. “Live” means that semantics are specified and the server behaves accordingly – for example, it checks client input or sets values automatically.
ILM for Experts: What’s the Interplay Between ILM and Traditional Data Archiving?

If you have implemented an SAP ERP 6.0 system and you are already using data archiving, then you are familiar with creating Archive Development Kit (ADK) files from object instances in your system during archiving sessions. During archiving, these files are always written to the file system first. Since SAP R/3 3.0, you have the option of moving the ADK files to a third-party storage system using the ArchiveLink interface. With ILM from SAP (based on SAP NetWeaver 7.0, enhancement pack 1) you have the additional option of moving your ADK files to storage using the ILM interface — that is, into an ILM-aware, WebDAV-enabled storage system.

In this case, IRM is used to determine the appropriate retention rule, and the ADK files are transferred via the XML DAS to the right place within the WebDAV hierarchy, along with the calculated expiration date, thereby enforcing the nondeletability of the ADK files. Because one ADK file contains many object instances that might not fit the same rule, you can split up and merge existing ADK files. You can also write files of a new archiving session to different WebDAV collections.

When you want to access archived data, ADK takes care of reading from XML DAS and WebDAV directly, without restoring the file into the file system. There are ILM-aware storage vendors that provide both interfaces, ArchiveLink and WebDAV, so that WebDAV appears as the ILM-aware alternative to ArchiveLink for accessing information on the same storage hardware. Such multi-protocol storage systems allow you to apply retention rules maintained in IRM also to ArchiveLink-stored unstructured data, such as images and scanned documents, even though these documents are not moved content-wise to WebDAV.

In more detail, the ILM-enhanced WebDAV interface not only requires the storage system to accept expiration dates and legal holds and propagate them in the hierarchy, it also specifies certain rules that have to be enforced to strengthen the retention setup from a compliance standpoint. For example, an ILM-aware storage system asserts and enforces that a retention period can only be prolonged and never shortened, or that an expiration date can never be updated to an earlier point in time. In this way, it guarantees the safety of the storage object well into the future.

Summary: What You Can Do Now to Prepare for ILM

A complete ILM solution is one that handles data from its creation through its final destruction. To help companies achieve this, SAP has been working on new ILM functions on the business application side — functions that focus on legal compliance and retention policies. SAP has also been working closely with storage partners to overcome the technological challenges of such a comprehensive solution. Through co-innovation with these partners, and by offering a certifiable ILM interface, SAP is giving its customers the freedom to choose from a variety of ILM-aware storage system vendors to be able to implement a complete ILM strategy.

8 Instead, the system creates proxy objects representing the ArchiveLink documents that carry the ILM properties. The storage system can take care of applying the ILM policy to the already stored unstructured documents so that they remain accessible to all ArchiveLink applications. They also inherit the enforced nondeletability.
So what can you do now to prepare for ILM? We leave you with seven important takeways:

1. Understand and upgrade to the driving technology behind ILM: SAP ERP 6.0 and SAP NetWeaver 7.0.
2. Put your retention requirements and corresponding policies on paper now.
3. Archive in adherence with your retention rules and policies; continue to use ADK.
4. Avoid a mix of data in archive sessions. For example, only archive data from the same fiscal year, or data with the same company code, together.
5. Use archive routing to segregate archived data based on archiving criteria in accordance with your legal compliance policies. This segregation facilitates a grouping of retained data similar to the nodes in the ILM hierarchy.
6. Avoid storing archive files on nonmagnetic media.
7. Ask your storage vendors if they will be supporting SAP’s ILM certification in 2008.

---

**Additional Resources...**

- “From Data Management to Information Lifecycle Management: Why (and What) You Need to Fill the Gap” by Dr. Bernhard Brinkmöller and Georg Fischer (*SAP Insider*, July-September 2007, [www.SAPinsideronline.com](http://www.SAPinsideronline.com))
- The Archiving your SAP Data seminar, offering comprehensive strategies and practices designed to satisfy your IT, business, legal, and audit requirements ([www.saparchivingseminar.com](http://www.saparchivingseminar.com))

---

---

* For more information on archive routing, see the white paper “ILM in an SAP Environment” available at [www.service.sap.com/ilm](http://www.service.sap.com/ilm).