



Information Lifecycle Management *Separating the Hype from Reality*

by Neil Murvin
Chief Technology Officer
CaminoSoft Corp.

Just behind fire and the wheel in terms of mankind's greatest inventions comes Information Lifecycle Management (ILM). Without a doubt, ILM will revolutionize every facet of the computing landscape. And along the way, it will probably generate more hype than all high-tech buzzwords of the past five years combined.

In this article, we put the concept of ILM under the microscope to determine whether it has true value. Does it solve a real problem? What does it encompass? How can the world of finance, education, government and the business world as a whole benefit from ILM? And how do organizations cut through exaggerated vendor claims to find the fundamentals that underlie it?

What's the Problem?

During the DotCom boom, companies sold a lot of hardware and software by trumpeting the latest wave of hype. They were, in many cases, selling solutions to problems that may not always have had a basis in fact. But they sure sold a lot of product.

A similar phenomenon is occurring today with ILM. Everybody is talking about it. All of a sudden, every vendor offers ILM solutions. The only way to keep your feet firmly planted on the ground, therefore, is to focus on your company's problems and your exact needs. What situations do you face in the world of storage, availability, archiving and backup that need to be addressed? What inefficiencies exist that, if remedied, would make a major difference to the business? Unfortunately, unless you have a genuine problem to address, no amount of

“ILM software” will offer any value. Fortunately, ILM is based on relatively sound, practical logic and can solve some pressing problems in the way that storage is utilized and allocated.

“Data is growing at 125 percent a year yet up to 80 percent of this data remains inactive in production systems where it cripples performance,” says Charlie Garry, senior program director at Meta Group. “To compound this problem, many enterprises are in the midst of compliance initiatives that require the retention of more data for longer periods of time, as well as consolidation projects that results in significant data growth.”

In healthcare, for example, organizations are struggling to meet the massive storage compliance mandates of HIPAA. Other legislation, such as Sarbanes-Oxley, SEC and state requirements, affects various other sectors including banking, government, higher education and business.

These diverse laws require strict record keeping and auditable records that must be stored for specific time periods and under specific conditions. The demands of storing corporate email alone, for example, could cause a tenfold expansion in storage needs over the next decade. Legacy architectures simply can’t be expected to cope with the projected demands – hence, ILM.

What is ILM?

Though varying significantly in its description from vendor to vendor, ILM is basically a strategy for policy-based management of information that provides a single view into all information assets. Spanning all types of applications and platforms, ILM continually aligns a given set of data with its storage resources – based on the value of that data to the organization at any specific point in time.

The Storage Networking Industry Association’s Data Management Forum has made substantial headway with standardization efforts, describing its vision for ILM this way: “Information Lifecycle Management is comprised of the policies, processes, practices and tools used to align the business value of information with the most appropriate and cost-effective IT infrastructure from the time information is conceived through its final disposition. Information is aligned with business requirements through management policies and service levels associated with applications, metadata and data.”

“The ILM buzz is similar to that surrounding virtualization,” said Steve Duplessie, an analyst with Enterprise Storage Group (ESG). “It encompasses cradle to the grave management whereby you get the right information to the right device or media at the right time.”

Does this sound a bit like Hierarchical Storage Management (HSM)? Yes, but HSM was historically one-dimensional. It typically involved one large server or mainframe and focused on objective measures like access frequency: if certain data hadn't been accessed in a specific time period, it was automatically moved to another type of media. ILM goes further with this same concept, extending it across the network to cover the entire infrastructure and adding subjective as well as objective criteria.

New regulations, for instance, largely negate the historic HSM time-stamp criteria. This philosophy was that if no one ever accessed a file, it would be either archived or deleted as it had indeterminate value. Such information often ended up off site stored somewhere in a vast tape vault. If you needed to retrieve it, an administrator had to physically rummage through the tapes in the hope of finding that one file in the tape vault haystack.

Under today's operating climate, however, that methodology is obsolete. Files that have not been accessed for years may now represent high value due to potential penalties that could be invoked if they have not been retained. Thus time-stamp, subjective and legislation-based data categorization are all incorporated into ILM. And regardless of where the data reside or the type of media employed, they can be managed from one console. Even with hundreds of applications, dozens of servers, terabytes of online and nearline data, and virtually unlimited offline archives, ILM is robust enough to cut through the complexity and manage an organization's information effectively.

Unbelievable Benchmarks

The whole concept of ILM, of course, suffers badly from its own over-hype. Excesses of the past are firmly ingrained in people's minds, so it's not surprising that IT professionals distrust vendors who claim that “ILM is the greatest thing since the PC.”

And this suspicion shows up in survey results. One CMP study revealed the following answers to the question, “Is ILM for real?”

- 23% feel it is marketing fluff.
- 17% say it is too nebulous to impact storage.

- 9% say it is just another acronym that means nothing.
- 30% feel it will be a big help with industry data retention regulations.
- 21% say it answers the need to integrate storage with business processes.

In summary, 49% of those surveyed are turned off by this new term. Therefore, let's take a closer look at the fundamentals that underlie ILM. Understanding the technology and how it relates to one's business provides an opportunity to assess whether ILM can truly add operational value.

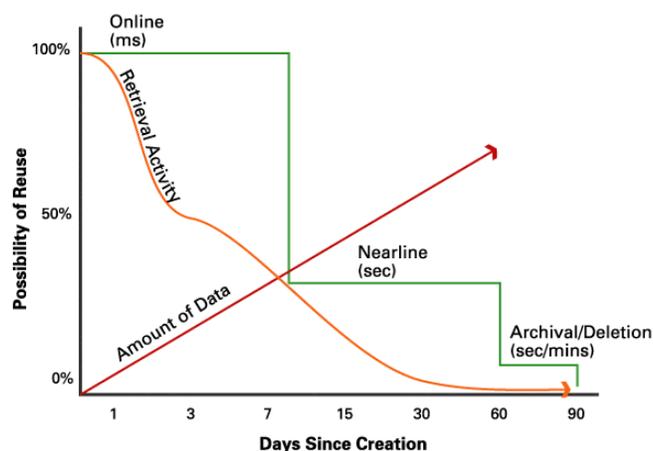
ILM Fundamental #1

Differentiate production data from reference data. Production data comprise those files that are actively utilized in day-to-day operations within the organization. Reference data, on the other hand, comprise those files that are not frequently accessed but still need to be maintained and available, whether in support of internal organizational requirements or due to external legal obligations.

ILM Fundamental #2

The likelihood of data reuse is directly related to age. Storage expert Fred Moore of Horison Information Strategies has studied data retrieval patterns, finding that access activity declines sharply over the first week following creation of a file (see chart); after one month, the information is rarely accessed. "The probability of reuse of data has historically been one of the most meaningful metrics for understanding optimal data placement," said Moore.

Data Reference Patterns



Source: Horison

ILM Fundamental #3

Provide structured control over file retention and deletion decisions. Let's say the example in the above figure is a bank. Transactional data is kept online only during its highly active period – the first seven days – and is accessible in milliseconds. Between seven days and two months, the bank moves the data to nearline storage, i.e., the information is kept on lower-cost disk arrays, retaining the benefits of transactional speed and high throughput. After two months, the bank then archives the data to offline media or deletes the data as mandated under its policies and industry regulations.

ILM Fundamental #4

Maintain compliance with mandated government regulations. In the healthcare field, for example, organizations cannot just automatically archive files or delete them based on a time-stamp. Current legislation calls for information to be retained and available for the duration of a patient's life and for several years afterwards. Merely pushing data to an offsite tape archive is insufficient. This calls for a flexible system that encompasses broad criteria other than "creation date," such as "file type," "includes/excludes," "last access date," "modified date," etc. Ideally, such a system will provide "hooks" for real-time, continuous measuring of data access patterns so that policy-based archiving can be adapted to "on-the-fly" administrator-supplied criteria.

ILM Fundamental #5

Maximize information availability and data protection. Servers, SANs, NAS devices, routers, switches and other critical components that provide the conduit between users and the information they need to do their jobs must employ sufficient redundancy to automatically "fail over" and recover from predictable faults in order to maintain information availability. Traditional backup alone – frequently performed "once-a-day" overnight and often causing information to become unavailable to users – is no longer sufficient to preserve business continuity. Service-level objective adaptation, performance tuning and retention control mean nothing if a disaster obliterates the unprotected information assets upon which they are built. We need to change the way we look at backup and redeployment utilizing today's live "snapshot" and replication technologies (local and remote) to provide greater protection, while maintaining continuous information availability to users and applications.

ILM Fundamental #6

Maintain application transparency of data for users. Organization executives don't have time to be concerned of data whereabouts. They simply want to query the system and obtain a fast response. Therefore, it is mandatory to provide a logical view of data that allows users to "see" files as they were created and in directories where they originally resided, regardless of their current location on different tiers of storage.

ILM Fundamental #7

Reclaim space on costly storage resources such as SAN. It is not uncommon for a business to spend over a half-million dollars on a SAN, so applications that minimize the initial need as well as future expansion are well worth considering. The objective is to maximize return on investment, which is accomplished by continuously monitoring and controlling information movement so that expensive SAN resources are occupied only by data that represent the highest possible value to the organization.

Summary

Information Lifecycle Management is all about ensuring organizational continuity while optimizing the deployment of IT resources.

In CaminoSoft's ILM suite, Managed Server HSM™ software provides an innovative, server-based solution that manages a wide range of storage solutions (SAN, RAID arrays, NAS appliances, general purpose servers, etc.). As an organization's storage requirements increase, additional resources can be easily added to the "Central Storage Pool," pinpointed to online, nearline or offline tiers on the basis of price/performance.

Storage management, however perceptive, is no substitute for structured backup, virus protection and server failover processes. For this reason, CaminoSoft's ILM suite is not only compatible with popular backup and anti-virus applications but also includes high availability solutions StandbyServer for NetWare and OFFSite Archive for NetWare. The former solution monitors the health and mirrors the volumes of multiple primary servers, providing both automatic failover and data access if storage becomes unavailable; the latter sends point-in-time "snapshots" of data to a safe remote server, facilitating centralized open-file backup, data replication and rapid disaster recovery.

Seamless tools that minimize operational stress – inherent in managing data, reclaiming storage resources, averting disasters, and avoiding cost – all work together to maximize efficiency. This is essential for the growth, viability and success of the organization.

About the Author

Neil Murvin is Chief Technology Officer at CaminoSoft Corp. For more information about Information Lifecycle Management, contact:

CaminoSoft Corp.
600 Hampshire Road, Suite 105
Westlake Village, CA 91361
Phone: (805) 370-3100
Fax: (805) 370-3200
Email: info@caminosoft.com
Web: www.caminosoft.com

© 2005 CaminoSoft Corp. CaminoSoft, the CaminoSoft logo and Managed Server HSM are trademarks of CaminoSoft Corp. All other trade names are the property of their respective owners.