Clipper Notes™

Practical Steps Toward
Information Lifecycle Management

Analyst: Michael Fisch

Management Summary

Taking care of valuable assets in our possession is part of smart living. It allows us to derive the most benefit from these things today and in the future. For instance:

- *If you owned a house, would you maintain the yard, roof, and interior to preserve its value and livability? Or would you let time and the elements take their toll?*

- *If you had a sum of money set aside for retirement, would you invest it to earn a return? Or would you hide it under a mattress?*

- *If you were the coach of a football team, would you require the athletes to train, show up for practice, and learn to work together on the field? Or would you let everyone take it easy and just hope for the best on game day?*

In each case, most would agree the first approach is wiser because it represents managing the asset actively and well, whether a house, retirement funds, or football team. It requires more work than the "do-nothing" option, but the benefit derived is significantly greater. Your reward is a house that you continue to enjoy, more funds for living later in life, and a team with greater cohesiveness, comradery, and chances of winning. Active management is worth the effort!

In our modern age of global, computer-driven commerce and instant communication, information is also an asset worth managing well. It is as worthwhile as people and money in the operations of an enterprise, since all of these are indispensable. Moreover, those businesses that manage information well will bolster their competitiveness, profitability, and longevity. Those that do it poorly will sacrifice these same things.

Information lifecycle management (ILM) is a strategy for managing data actively, thoughtfully, and according to the long-term interests of an enterprise. Its objective is to deliver the right service levels to the right data at the lowest cost. ILM is both practical and visionary, and is based on several core principles of information and management.

Moreover, there is a common process to follow in implementing it, even though each enterprise's results be different, reflecting its unique business and data requirements. It is time to get started! Read on for details.

Foundations of ILM

Much has been said and written about Information Lifecycle Management (ILM). People have different conceptions and understandings of it, making it a tricky subject to write about. One should begin with a definition, and here it is: *ILM is a strategy for managing data actively, thoughtfully, and according to the long-term interests of an enterprise.* It is not a far-out or ethereal concept. ILM simply states that you should be smart about managing data because it is a valuable asset.

Even so, some people question the utility of ILM because it presents a vision for managing data. These are often pragmatists who have a predilection for action, results, and systems that work now. To them, ILM might seem too lofty and conceptual. However, this opinion is probably the result of a misunderstanding, because ILM does not have to be done perfectly to provide a valuable service. In fact, **the true spirit of ILM is to take practical, incremental steps toward better information management.** In this context, possessing a vision is good. It means you have a positive direction for progress.

ILM is based on several core principles of information and management, which follow.

Business Determines the Value of Data

Businesses use and generate information every day. It comes in many forms: invoices, employee records, e-mail, marketing materials, and so forth. These waves of information accumulate and grow over time. It is important to store, manage, and protect the data properly, because it is necessary for daily operations. A business needs the data to run.

Moreover, each business determines the value of its data. No two organizations are alike, nor do they contain the same data. Unlike the situation of a $20 bill, which everyone would agree is worth twenty dollars, the value of data is relative to the business that owns and uses it. Consider the example of client records from five years ago. A medical clinic would say they are quite valuable because laws demand their retention and they provide historical context for diagnosis and treatment. A dry cleaning shop, on the other hand, would have almost no use for client records of that age. Not worth the expense to keep them, the shop would probably toss them out. In short, **data serves the interests of a business and derives its value from that business context.**

The Value of Data Changes

Like a tree in a forest, a data artifact follows a cradle-to-grave lifecycle. It initially springs to life as the creation of an application and user. This data could be of high value, such as a recent customer order, or of lesser value, such as a departmental e-mail about a pick-up basketball game after work. Who or what created the data, the subject matter, the related business processes, and other factors determine value. Furthermore, data often begins in an *active* state, where it is accessed and altered frequently, as in a document undergoing an editing process. After a period of time, it can become *fixed* or *inactive*, where it is much less likely to be accessed or changed. Fixed content still needs to be retained for business and regulatory purposes, but slower access and recovery times would be acceptable. An unexpected event, like a financial audit or a long-dormant customer deciding to repurchase, could revert data back to an active state. After a (possibly very long) period of time, the value tends to fall to zero and data can be deleted. This ends the information lifecycle, whose motto is: *Not all data is created equal, and values change with time.*

Tiered Infrastructure is the Foundation

Tiered infrastructure is the foundation for treating data differently based on its relative value. It can be likened to sitting down in a restaurant and choosing between filet mignon, a steak, and a hamburger. The difference between them is a matter of quality (of taste) and cost. Filet mignon is the most flavorful and expensive, while a hamburger is affordable on a college budget.

Similarly, tiered infrastructure includes the storage systems, networks, and data management software that an enterprise uses for storing, protecting, and securing its information. It is segmented into different qualities of service – performance, availability, recoverability, security, retention periods, and cost. The tiers might consist of different types of storage systems, disk drives, RAID configurations, data protection and replication software, security applications, and archiving.

Service Levels Are Key

In the figure below (*The Arch of Information Lifecycle Management*), data and tiered infrastructure are shown as part of an arch supporting the business. The keystone of the arch is *service levels*, because they are the connection and common ground between data and tiered infrastructure. Service levels represent the consumer- or demand-side view. This is like the difference between a power plant and the electricity service delivered to your home. As consumers, we appreciate and experience the service delivered to our doorstep. The infrastructure that provides it, including all of the points of manufacturing and distribution, are behind the scenes and taken for granted. In a similar vein, storage service levels are what the application and user actually experience. They are defined by specific metrics of performance, availability, recoverability, security, etc.

In this mode, data is classified and mapped to service levels based on application, user, and business requirements. Then, tiered infrastructure is deployed to meet these service level requirements. In this way, the arch of ILM is completed.

### Right Service Levels to the Right Data at the Lowest Cost

What is the specific objective of ILM, beyond generally managing data well? The answer is **to deliver the right service levels to the right data at the lowest cost.** The balance is walking the line between the extremes of under-performing and overspending. To continue the dining analogy, you would not want to offer hamburgers if you are hosting, for instance, the Queen of England. However, if you are catering a children's birthday party, why pay for filet mignon when everyone would be happy with hamburgers?

Productivity and profitability may be affected if you do not provide sufficient storage service levels for your enterprise. If a database requires a certain response time to keep up with customer transactions, then the performance should be available. If regulatory compliance says data should be retained for a certain number of years, then archive it. If the enterprise wants to recover quickly from a local disaster, then replicate to a remote failover site. This is all part of delivering the right service levels to the right data.

Since storage and data management are a sizable component of the IT budget, it is also important to classify data and place it in the most cost-effective tier. ILM offers a perfect opportunity to take advantage of less-costly storage technologies – such as SATA drives, parity RAID, and IP network connections – as well as to proactively delete data when it is no longer needed and/or becomes a liability.

Exhibit 1 – The Arch of
Information Lifecycle Management
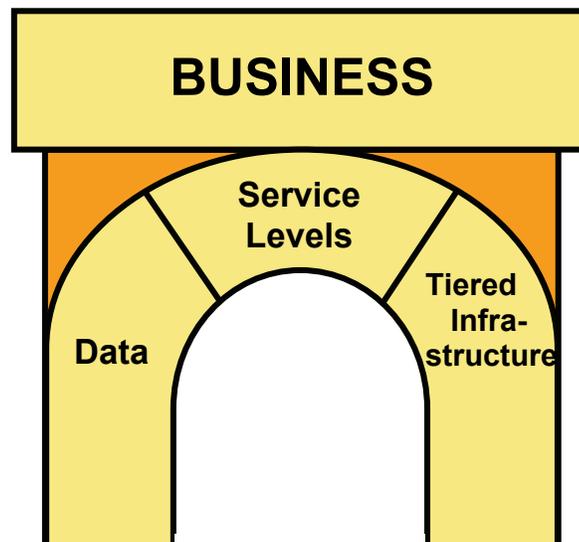
Practical Steps Toward ILM

There is a general process to follow for implementing ILM, as embodied in these four steps.

1. Discover your information assets.
2. Classify data by required service levels.
3. Map service levels to infrastructure tiers.
4. Actively manage data over its lifecycle.

Like the data itself, the way each enterprise implements ILM will be unique. Some may take a comprehensive, in-depth approach, while others may opt for something basic and yet effective. It depends on what your business needs, and it will evolve with time.

Discover Information Assets

The starting point to ILM is discovering your information assets, because it is necessary to know what is there before you can manage it properly. You might call this *taking inventory of the data*. This process is like walking around a virtual warehouse to itemize its contents, which includes quantities and item descriptions (i.e., metadata).

Data assets include records contained in *databases* (structured), *messages, and attachments* in e-mail applications (semi-structured), and *files* (unstructured). Files generally comprise the largest portion of enterprise data. It is also the most challenging to inventory because of the sheer number of files (potentially into the millions or billions), variety of formats, and the descriptive limitations of standard file attributes (name, type, date modified, etc.).

Data discovery can be manual or automated. The manual approach is to look at each application and file server in order to inventory and characterize its associated data. This is labor-intensive and of limited descriptiveness, though feasible in smaller environments. The alternative is to use automated software tools. Storage resource management (SRM) software scans the environment and provides a synopsis of databases, file systems, and their associated volumes on storage arrays. File information can include summaries based on file attributes, such as the total number of Word documents or files that have not been accessed for over a year. Even more advanced file classification tools are now available that scan and report on file attributes *and contents*. It is like opening and reading the letters instead of just looking at envelopes. These tools enable sophisticated classification according to specific business rules, as well as full-content searching for legal discovery and general business purposes.

Classify by Required Service Levels

The next step is to classify data according to service level requirements. Begin by sorting data into meaningful groups that have similar requirements. This may be done at the level of discrete applications and file systems (e.g., databases, e-mail, file servers) or more granular data subsets for lifecycle management. Metadata and content indices gathered during the discovery are especially helpful for classification. Next, specify the service quality each data class should have, in light of business needs and cost tradeoffs. The service levels will reflect access performance, availability, recoverability (RPO and RTO), security, archiving, and retention period. In addition, there will be a set of policies associated with each service level that specifies how to put them into effect. For instance, the policy to provide a high service level for mission-critical, transaction-processing data might be: high-end storage, Fibre Channel drives, RAID 1, redundant SAN connections, data replicated to a remote site for disaster recovery, periodic snapshot copies, non-disruptive backup, limited user access for security, archiving records to low-cost disk after one year, and deleting records after seven years.

Classification is probably the most challenging aspect of ILM because it presents more than a technical problem – it requires an organization to make subjective judgments about the value of its data. This should be a collective effort, because no single stakeholder has a full sense of this issue. Functional and departmental personnel have the best understanding of their particular business process and application requirements. IT personnel have the best understanding of technical issues and ramifications. Executives control the purse strings. Where regulatory and legal requirements are concerned, executive, legal counsel and security officers will want to provide input and direction. Communication and a consensus-oriented approach are recommended to ensure service levels are both realistic and reflect the needs of the whole business. Departmental chargeback is a good way to monetize the value and cost of service

levels, which encourages accountability and realistic choices in resource consumption.

Map Service Levels to Infrastructure Tiers

If service levels express how an enterprise *intends* to treat its data, tiered infrastructure represents the *reality*. This includes the storage hardware, software, and processes applied to data in an effort to achieve the specified service levels. Therefore, an enterprise must map its service levels and data onto a tiered infrastructure.

A disk array and tape library would be the most basic tiered storage infrastructure, but technologies today allow many more gradations of performance and price. Low-cost disk arrays with SATA drives are especially effective for second-tier, online storage, such as archiving, disk backup, and testing and development. Snapshot copies, remote replication, and continuous data protection (CDP) can be applied to tiers based on data protection and recovery needs. Though the possibilities are endless, deploy a number of tiers (with associated hardware, software, and processes) that allow for a balance between effective service level delivery and saving costs. Beyond a certain point, there will be diminishing returns for additional tiers because of management overhead. A rule of thumb is three to five tiers, but it really depends on your environment and capabilities.

After completing discovery and classification, you may find there are more service levels than storage tiers in your infrastructure. That is okay, because you can map multiple service levels (and data classes) into a single tier. The idea is to "round up" – that is, meet or exceed all service level requirements in a tier. Technology innovations in the future may let your enterprise deploy additional tiers and achieve greater precision in service delivery.

Actively Manage Data Over Its Lifecycle

The last step is an ongoing one – active management of data over its lifecycle. An important aspect is taking care of the data itself, which starts with the initial placement of data into the tiers. You can then migrate data periodically between tiers for purposes of archiving and lifecycle management – and eventually delete it. Manual migration using traditional techniques is possible, but it requires ongoing effort and downtime. Alternatively, products are available now that perform easy,

non-disruptive migration in specific environments, such as for file systems, volumes in storage systems, or particular applications. Some can also migrate data automatically based on rules while retaining transparent user access to data (i.e., policy-based archiving). These deliver the highest value of ILM, in terms of cost savings and improved performance and availability of the primary application. In addition, tools are available for modeling, analysis, and reporting. They help with tasks like capacity planning, finding, and eliminating redundant files, understanding data access patterns, and projecting the ramifications of taking certain actions.

The other management aspect is leveraging data to make the business more productive. For instance, the ability to perform content-based searches is hugely beneficial. Imagine being able to handle a legal discovery in an hour, as opposed to days or weeks. It saves time and expense and lowers risks. Repurposing data for business analytics or application testing and development is another useful capability. The possibilities are manifold.

Conclusion

Once again, ILM is a matter of managing data actively, thoughtfully, and according to the long-term interests of an enterprise. Like investing money or coaching a team, there are many different ways to do it, and the important thing is active engagement. **It is not a matter of perfection, but taking practical, incremental steps – sooner rather than later – because information is an asset worth managing well.**

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

➢ *The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.*

About the Author

Michael Fisch is Director of Storage and Networking for The Clipper Group. He brings over ten years of experience in the computer industry working in sales, market analysis and positioning, and engineering. Mr. Fisch worked at EMC Corporation as a marketing program manager focused on service providers and as a competitive market analyst. Before that, he worked in international channel development, manufacturing, and technical support at Extended Systems, Inc. Mr. Fisch earned an MBA from Babson College and a Bachelor's degree in electrical engineering from the University of Idaho.

➢ *Reach Michael Fisch via e-mail at mike.fisch@clipper.com or at 781-235-0085 Ext. 211. (Please dial "211" when you hear the automated attendant.)*

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.