Clipper Notes™

VENDOR-AGNOSTIC EXPLANATIONS AND ADVICE
FOR THE INFORMATION TECHNOLOGY BUYER

Navigating Information Technology Horizons

ILM Stage 2 —
Full Spectrum Information Lifecycle Management

Analyst: Anne MacFarland

Management Summary

Increasing re-use of business information makes Information Lifecycle Management (ILM) more than just a storage issue – and more than just an IT issue. ILM goes beyond distributing information in library systems or even sharing and transforming information in the formalities of the research process (both paradigms of the past). Now, it addresses using business information to create advantage and to sustain and optimize operations. These uses pertain both to managing a data center and to running a business. It is the dawn of ILM, Stage 2.

Consider a real world example of enhanced use of information. The New England Patriots, a professional football team, make extensive use of the game tapes that are owned by the league in which they play and are freely available to all member teams. It is what the Patriots do with the video that makes all the difference. Each member of the offense can focus on the reactions to different strategies and play options of the players that will cover him. With the unique intelligence he brings to the situation, he can find the patterns that will let him out-think and out-play this opponent in the coming game. This member of the offense can also see clips of the broader defensive strategies of the coming opponent and can craft plays to leverage the opponent's propensities. Of course, there is also training, practice, conversations, and more group film viewing to build the broader strategy for the next game. This is a different kind of preparation than getting a good night's sleep, eating right, and making sure your equipment is in good repair.

The ability to present information at various granularities to meet immediate needs is the basis of the huge difference between what we can do now and all the prior modes of information transfer. We have the information tools to start nibbling at the *more* that can be done with information at many levels, and many new possibilities are coming to market. This is not a simple push from a data center; considerable input from users is involved. The user's intelligence and particular focus determines the sources, variety, and granularity of information that is needed – and these needs will change frequently. With it, business people, just like the New England Patriots, can use their intelligence, together with well-targeted information, to address the operational challenges that they face, be it management of a process, development of products and services, or the business of selling. It then falls to IT to make the needed information findable and usable – not as incomprehensible query results (think of the ancestral green screens and weighty printouts) ranked by some general-purpose concept of relevance but as a highly-specific deliverable, often with real-time value. At the same time, IT must continue to maintain the integrity, security, and privacy of the information.

Much could be done, if money and time were unlimited. In most businesses, they are not. **It is best to focus on using the full spectrum of information lifecycle management to mitigate business risk and enhancing opportunity.** For more on why you want ILM Phase 2 tools, what they are, and how they work read on.

The Impetus for a More Comprehensive ILM

The *least-cost, good-enough* of commoditization has become a brittle basis on which to build a business. In a global economy, hordes of competitors can and will undercut some aspect of your product or service and put you in defensive mode on many fronts. For sustainable success, you must find a unique *more* that you can provide. **Lasting value – and what cannot be imitated – comes from what you can do better based on all the wisdom you have learned from your time in business (or, in many cases, in a series of businesses) together with the strategies for using information that you have also honed over time.** This wisdom gives insight into your processes – and how to evolve them safely to meet new situations. The information on which the wisdom acts comprises more than just the lists of customers and what you have sold them. It includes the knowledge of their experience of your products. It includes your knowledge of what kinds of enhancements you can give your products – and how fast. All this information is locked up in your corporate data. **If you only see corporate data as capacities on a disk or tape, it is hard to appreciate – to seize on, and to leverage – its value.**
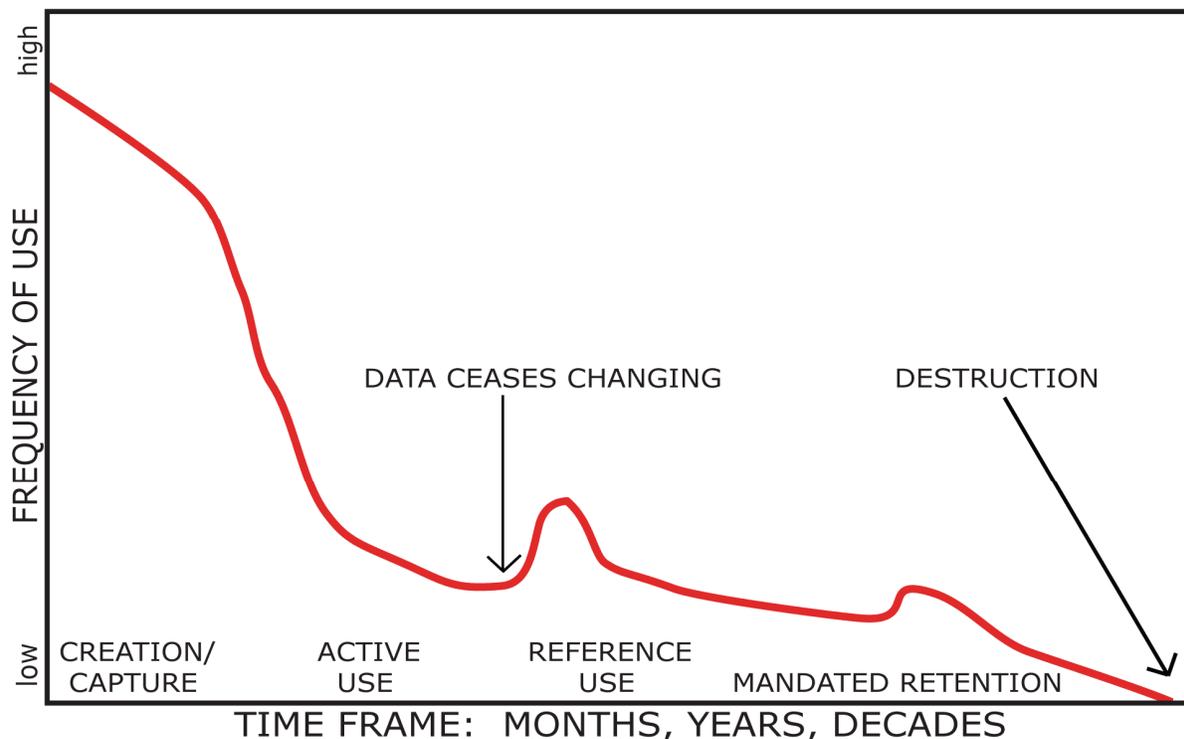
Stage 1 ILM

Storage vendors coined *Information Lifecycle Management* and its variant *Data Lifecycle Management* as strategies to mitigate the burgeoning cost of required storage. Factors contributing to the high rate of growth include:

1. The regulatory compliance requirements to keep more documentation of business activities,
2. The business need to keep more information about customers with whom they increasingly have had no face-to-face or persistent relationship.
3. The lack of aggressive deletion strategies due to uncertainty about the direction of business model and regulatory change.

ILM Stage 1 **leveraged the waning use of aging information to promote storage cost savings via tiered storage environments.** Information that was needed less frequently and less urgently could be stored on disk arrays and tape libraries that could meet the service levels required at more moderate cost. Migration strategies and "transparent migration" products were developed to meet this need, and stubs were developed to keep IT systems unaware of any bifurcation of the physical location of the data.

Exhibit 1 - Information Lifecycle, Stage 1

FREQUENCY OF USE (high / low)

DATA CEASES CHANGING   DESTRUCTION

CREATION/CAPTURE   ACTIVE USE   REFERENCE USE   MANDATED RETENTION

TIME FRAME: MONTHS, YEARS, DECADES

A traditional ILM chart is shown in Exhibit 1, on the previous page. Note that, with the growth of Web–based commerce, the end-of-business day (or week, or quarter) statistical accumulations on which reporting was based have been supplanted by more granular hourly statistics. The point at which information becomes documentation (ceases changing) is moving inexorably to the left of the chart. At the same time, mandated retention of business data has extended the lifecycle. This exposes a growing expanse of static information about the business, its rhythms, and its characteristics that is available to aid business decisions.

New Business Challenges

As enterprises have become more physically distributed and more prone to base their evolution strategies on mergers and acquisitions, the emphasis in ILM has shifted firmly from *Lifecycle* to *Information*. Business challenges abound.

- **Keeping customer and product information consistent across multiple locations** has become both difficult and a priority. Tools like Master Data Management and caching strategies address this.
- **Keeping people's awareness synchronized across these multiple locations** is difficult, so we have the aggregations of global/wide area file systems, collaborative spaces, and various publish and subscribe push mechanisms. All these have significant security consequences.

- **There is a dearth of ways to determine the overall health of today's multi-faceted, highly-partnered, somewhat-outsourced enterprises**. Building this baseline is partly a matter of portal-based dashboards and well-targeted collaboration, but increasingly, it is most efficiently addressed by growing and better using the institutional knowledge base.

Uses of Business Information

Up-to-the-minute news and accurate real-time business metrics are part of business-as-usual. Over reliance on them, however, can give a curiously flat perspective. To address unexpected opportunities properly, an institutional knowledge base needs a greater depth of focus and a more comprehensive view than real-time information gives. More specifically, if a similar opportunity worked out badly a few years ago, it is good to be able to analyze the factors involved in the failure.

The knowledge base must also support the frequent need for *all* the relevant information in order to assess strategy and to provide a rich *variety* of information with which to make decisions. The identification of key performance indicators (KPI) is a wonderful tool for keeping people focused on certain kinds of productivity, but the KPI paradigm fits poorly with matters of research and investigation. Therefore, the more kinds of relevant information that can be brought to bear, the better the outcome will be.

Exhibit 2 - Information Lifecycle, Stage 2
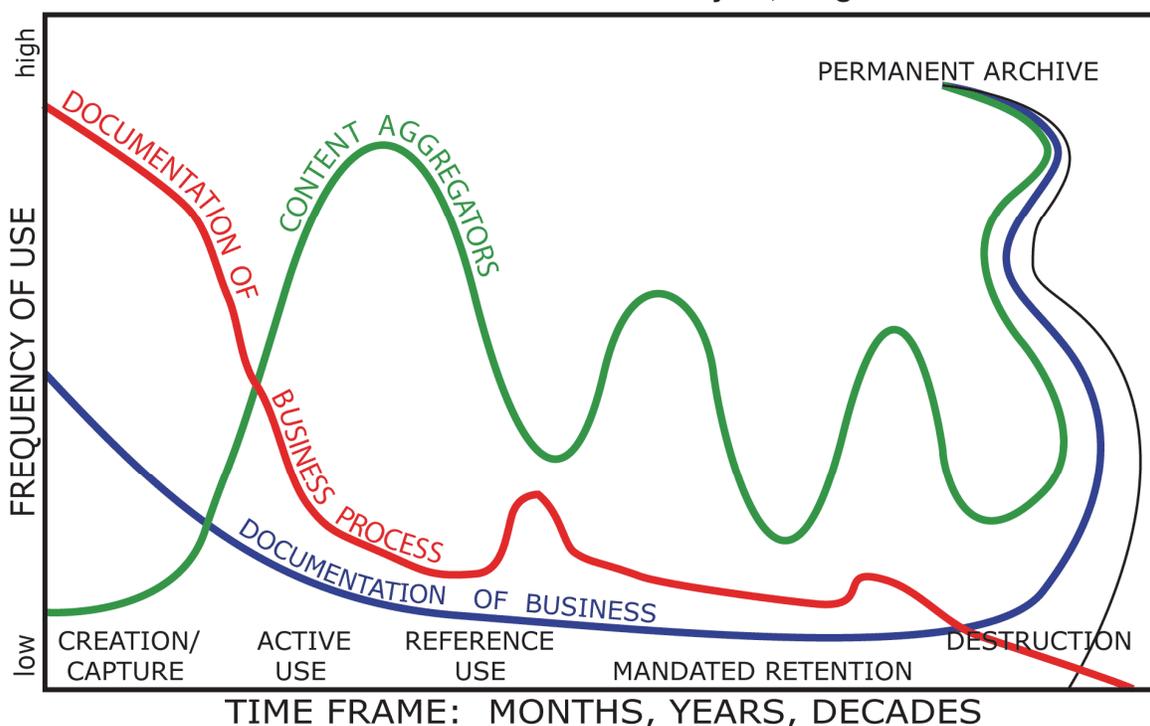
And, there is a lot of variety to leverage. The information from IT instrumentation and sensors, like RFID tags, gives a wealth of information about processes, people, products, and other assets and their logistics, much of which is of critical-but-transitory importance. Such information, taken together, also gives a different view of process than the workflow or report. Mashed up with maps, it can surface snags in process that a top-down view would not find.

These kinds of business information are not always reusable without aggregation, dissection, transformation, or enhancement. Business intelligence applications and text analytics are examples of these information operations. Think of how such information functions could make other corporate information more reusable.

As seen in Exhibit 2on the previous page, once an organization starts leveraging the tools of Full Spectrum ILM, the content aggregation and reuse will change the profile of information use over time from that of ILM Stage 1. Long-term or permanent archives of information are a way to keep information well organized for chronic use.

The Tools of Stage 2 (Full Spectrum) ILM

For any use of business information, the need for security and data protection is obvious. Data quality checks are also important. Without quality information, there are no quality business strategies. Gaps in documentation decimate the ability to model what-ifs, so proper capture of paper-based business data is often a key part of a comprehensive ILM strategy. Information services, be they classification to a taxonomy, indexing for key words, entity extraction, or relational analysis of structured and unstructured data (sometimes called data profiling), are all valuable but expensive. How money is appropriately spent comes down to a weighing of both risk and reward, considered from the business point of view as well as that of the IT system.

ILM Stage 1

The first kind of tool of any ILM Stage 2 initiative will always focus on the *storage and infrastructure* of ILM Stage 1 that supports the use of information throughout the enterprise. Without system level security, and without the basic data services of data security and resilience (comprising access control, replication, and disaster recovery), you cannot build the high-value structures and tools that optimize data use. If your structured data is not cleansed and if your unstructured data is stranded in various laptops or laboratory equipment, your ambitions for better use of business

information will be limited. Part of establishing an environment for the use of information as a strategic valuable asset requires some use of WORM[1] and encryption.

Metadata and Other Characterization Tools

The second tool of optimized information use is *metadata*. Metadata, information about data, has been around for many years in various forms. Metadata is also a key part of relational databases, forms, and other data structures.

Its current incarnation is as XML and its variants, but it is not a new concept. It has the advantage of being an open standard, though sometimes it is honored in the breadth by proprietary extensions.[2] Metadata aggregates around data in the form of a time stamp from the moment of its capture. IT processes generate metadata, as do presentations and, if you use Information Rights Management, usage metadata. Metadata documents data attributes, dependencies, and other relationships. Several years of work have been in process to incorporate standard metadata elements in standardized objects for particular industries and for educational domains. More general standards for unstructured information, such as open-source UIMA[3], and for ontologies, such as OWL[4], are also part of making metadata a more effective information management tool.

Metadata as a tool underlies indexing, search, filters, and analytic tools. Presentation metadata is key to putting information into a (literal) frame of reference that is relevant to the user. Think of metadata as sugar, a spoonful of which will help the medicine go down.

Classification and Other Aggregative Tools

The third class of tools is *containers* - objects, repositories, classification schemes, and other collective devices - that foster manageability and understanding for an aggregate. Like metadata, containers – and objects in particular – have been

---

[1] WORM (write-once-read-many) prevents overwriting or changing information. It is found in many forms of media including disk, tape, and optical.

[2] Such extensions can play a key part in keeping standards properly evolved to address new challenges, if they become part of new standards.

[3] UIMA (Unstructured Information Management Architecture), like many of the specific elements mentioned in this bulletin, is not a product but a standards-based tool. Developed at IBM, it has been open sourced. Information about UIMA is available at http://www.uima-framework.sourceforge.com.

[4] OWL (Web Ontology Language) was accepted as a standard by the W3C.org in 2004. A multimedia OWL has also been developed.

the subject of lots of standards work and are worth a fuller explanation than is possible in this bulletin. While policies may be written to particular metadata elements, they are more likely to be written in terms of a domain, such as a repository, or a container, such as a data set.

Presentation

The fourth tool category, one somewhat alien to data center traditions, is presentation. If information is not presented in the right place at the right time and in a readily consumable form, its value is greatly diminished. It is easy to overlook what is often a classic last mile problem and sometimes hard to understand that access to all the information in the world does not make you smarter until you can use it easily. Content management systems and portals are typical and useful presentation tools.

Recently developed lightweight development tools, such as PHP and Ruby, provide a more develop-on-demand capability. Environments such as AJAX (Asynchronous Java and XML) allow faster, less intrusive updating of information in a browser. These tools enhance how data centers can present information to their users.

Analytics, Visualization, and Other Use-Focused Capabilities

The fifth set of tools that pertain to the full spectrum of ILM are the *analytics and visualization applications* that are key to inspiring insight and revealing knowledge to users. Known for years as *Business Intelligence* and *Knowledge Management* tools, they are now available in various strengths to match user expertise.

How you use these toolsets will depend on how you use information in your business. If your business revolves around logistics (shipping, plumbing, etc.), your presentation mode may be skewed to PDAs, cell phones, and other small devices. This has limited the choice of visualization and analytic tools in the past, but most software vendors see the huge opportunity and are designing their applications for presentation on small devices.

Exhibit 3 - A Recipe for Organizational Functionality

Prerequisite: IT with good practices of security, data protection, and disaster recovery.

Step 1: Have your business and IT people identify sources of risk (e-mail, information shared with partners, and islands of inconsistent information are examples) **and opportunity** (research, successful initiatives, feedback from customers might be examples). Determine how these kinds of information can be used more effectively and, if appropriate, leveraged across the enterprise. These are the equivalent of fruits in season – the things on which to focus.

Step 2: Survey your IT assets (Intranet, shareable information, collaboration spaces) **to determine if your infrastructure will support your information use ambitions.** This is like checking your pantry and tools.

Step 3: Prepare your ingredients. Data may have to be assessed for quality, transformed, staged, and/or federated. Some file systems will have to be pared to expose their contents for indexing.

Leverage the commonalities of XML (keep usage consistent) to expose metadata available from document properties, database structures, file systems, and email attributes to derive elements of *provenance* (where the information came from), *relevance* (what the scope of the information is) and *context* (what larger information structure and related structures pertain). Extract content-based attributes, as desired.

Step 4. Prepare information to suit, using tools such as taxonomies, indexing, watermarks, repositories, database structures, and the hardware tools or striping (parallelization) then cache, dice, and otherwise containerize your information to minimize contention from concurrent use, or offload for greater process efficiency.

Step 5: Provide your clientele with place settings of search and analytic tools, and a presentation mode tailored to their needs (portal, dashboard, widget, etc.).

Business Use of Full Spectrum ILM

Use of these tools is summarized in the recipe card shown in Exhibit 3, at the bottom of the previous page. Like any good recipe, the specifics are more suggestive than prescriptive – but the order of the steps is important. If you don't brown the onions first, a stew just does not taste the same. If you don't secure your infrastructure and survey your assets accurately before you build your information value-adds, your reuse strategy will be hobbled by unexpected snags.

Seen as a recipe, the first stage of ILM is like the parental doling out of macaroni and cheese, where the choice was of infrastructure (microwave or range-top), and the tradeoff was higher cost (microwave) versus more manual things to do (pot scrubbing). By contrast, *Full Spectrum ILM* is a fusion-cuisine short-order diner/gourmet restaurant, where the menu is extensive and additional offerings are available to good customers.

Full Spectrum ILM for the IT Environment

Another way to grasp the capabilities of what Full Spectrum ILM can do for you is to look at the familiar domain of IT management. In the early days of IT, dials and lights had to be kept in the safe zone and green, respectively. This information was persisted not as bits and bytes but as logs that were the source for redress of faults, not for active management. With networking, the management information stream became digital. Information could be presented in a browser. It could be trapped and filtered to alert for particular conditions. It could be modeled to test out changes before the change was implemented. It could be profiled to produce the characteristics of normal at a given time of day or day of the week. As a result, IT management became more proactive.

From a more business-facing IT perspective, information becomes a way to see how business process was done, but this function was often seen as a by-product, not an asset in its own right. Reports were the deliverable. The information that was the basis of those reports was just a capacity to be managed and persisted through backups, which you hoped went well. Things have changed. Memory and storage are less expensive, so a larger set of historic information can be saved. Governmental regulations require that publicly held companies and companies that pay taxes or do business with the government retain their business records well and produce them on demand. Not using the information you are required to keep for business advantage is a wasted opportunity.

Conclusion

The physicality of where and how your business information is stored is critical to its proper access, but that alone will not insure its full and proper use. In a competitive world that is rapidly changing, how well your business can access its information and leverage the information and processes that differentiate it from the competition is critical to organizational longevity. The bitter winds of consolidation are howling throughout the realms of business, and the prospects for a business focused merely on survival are bleak. **By embracing the full spectrum of Information Lifecycle Management, you can know more, be wiser, and prosper.**

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

Anne MacFarland is Director of Data Strategies and Information Solutions for The Clipper Group. Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

➢ *Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 128. (Please dial "128" 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

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.