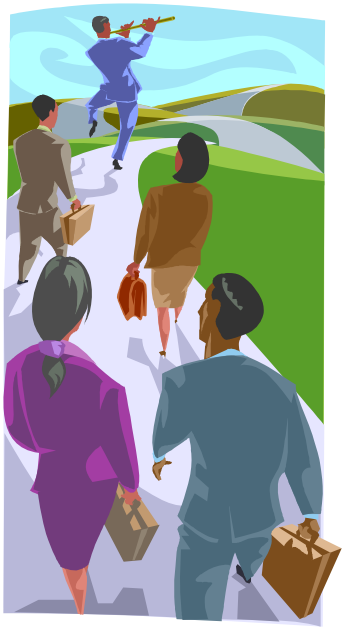
A report for the Society for Information Management
Advanced Practices Council
Driving Competitive Strategy Through Thought Leadership

Information Lifecycle Management Concepts, Practices, and Value



James E. Short
University of California, San Diego

August 2007

SIM
Delivering Business Value Through IT Leadership

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Driven by new government regulations in corporate disclosure, records retention and security, Information Lifecycle Management (ILM) became a storage and IT industry buzzword, both trumpeted and criticized for its promise of aligning the business value of information with the most cost effective and flexible IT needed to provide it. An approach to information management that seeks to take into account information's changing value, ILM is a process for managing information through its lifecycle.

This report summarizes findings from a year-long study undertaken to explore the views of senior technology managers in addressing ILM. Executive and professional surveys in fifteen case study organizations and an online survey of 345 IT and storage managers were completed. These interviews sought to understand how technology executives viewed ILM, its business drivers, its implementation record to date, and its ultimate business value.

The main drivers of ILM are: enterprise data growth; growth in unstructured data; limitations in relational data base management system performance; information access and security concerns; lack of effective methods for classifying data; and difficulty in assessing productivity of systems, applications and databases.

The main benefits cited of ILM are increased control over data, regulatory compliance (thereby minimizing business risk), and reduced costs (by eliminating redundancies in data storage). However, half the respondents felt that to reap these benefits, systems and organizational changes must be instituted. Respondents also pointed out the key challenges in instituting ILM: high cost and complexity of the storage management environment; lack of standards, leading to confusion in the marketplace; and required up-front investments in data, applications and storage hardware.

The report cites twelve lessons from early ILM initiatives:

1. **Focus on information value and the processes used to extract value**. Technological factors have made it easier and more efficient in the short term to use technical means to store increasing amounts of information. Management factors such as business risk, compliance, and business continuity of sorted information now make it desirable for business and IT to address the issues of valuing information from its collection and use, to storage, and to the time it is discarded.

2. **Focus as much on discarding information as storing it**. The majority of companies in the study had inconsistent (and incompatible) policies for discarding information. Moreover, there were questions regarding the effectiveness of policies to discard data, given the practical realities of data replication, multiple copies, backup and recovery copies, archival copies, and increasing data stores in business applications on enterprise and departmental networks.

3. **Initiate ILM through policy and people, not technology**. ILM is a management process implemented through policies, by people, supported with technology, in that order. Companies were skeptical of the value of technology initiatives proceeding under an ILM banner without clear business sponsorship and active line manager involvement and support.

4. **Recognize that the primary drivers of ILM are compliance, legal discovery, risk management and data retention**. The main drivers for initiating ILM projects were compliance, legal discovery, risk assessments for archived information, and security and regulatory requirements for data retention. These drivers have led to a proliferation of tactical responses and a focus on email as the application with which to start.

5. **Acknowledge that email's value is user, not enterprise, defined.** Email presents the difficult problem that value is defined in the user's terms, not the enterprise's. And user studies of email indicate that users are very reluctant to discard old emails because they are not only a means of communication, but also a storage resource for archiving project work.

6. **Recognize that application-specific ILM implementations can disrupt enterprise-wide initiatives.** The vision of ILM may be enterprise-wide and holistic, but firms are deploying application-specific solutions in response to short-term needs, such as complying with data retention regulations or improving storage utilization. Given the complexity of any scaled-up ILM initiative, this pattern will hold for some time.

7. **Implement storage best practices.** ILM initiatives usually start with a tiered storage architecture, where mission critical transactional data are stored on high performance disk systems attached to servers (online storage). Less critical data, such as months-old sales or inventory data, might be stored on a storage area network made up of less costly, slower drives. And document back-ups from enterprise PCs might be stored on archival tape systems.

8. **Define lifecycle classification and data movement processes.** Newer classification software is designed according to four principles of data access, use and migration:

    - Access and initial placement define the performance characteristics seen by the user in the application.
    - Recovery and protection define what happens in the event of a primary data failure.
    - Discovery, retention and disposal define the service characteristics of archived data.
    - Security and access control define how data is protected from unauthorized use.

9. **Lead technically with assessment of search and automated classification software.** Two of the most important technical and policy aspects of ILM are search and data classification. Improvement in these technologies is necessary to advance the problem of defining information value, especially more complex multimedia data.

10. **Review e-Discovery preparedness**. New U.S. Supreme Court rules for electronic discovery of documents in civil cases took effect in December 2006. These rules specify requirements for submitting electronic documents, including email and other forms of communication, as evidence in civil cases.

11. **Assess the business case for ILM**. ILM's promise is to help organizations better manage their data (and therefore the IT infrastructure that manages the data) from the time that data is created until it is no longer needed. On the hardware side, tiered storage enables different services and service levels, offering potentially different cost structures. On the process side, data is characterized to better describe its specific value to the application and business process. Software infrastructure enables reporting, protection, and data migration capabilities. Only when all of these are available can ILM be fully implemented and its benefits realized.

12. **Recognize that there is no one ILM blueprint, although a direction is evolving**. Firms have adopted multiple approaches, depending on the scope, objectives, and level of sponsorship of ILM projects. Enterprise level ILM and data warehousing efforts require executive level commitment and stewardship across the full planning and project lifecycle because they involve customer data and the information necessary to meet functional or enterprise goals.

I. Introduction

Driven by new government regulations in corporate disclosure, records retention and security, Information Lifecycle Management (ILM) became a storage and IT industry buzzword, both trumpeted and criticized for its promise of aligning the business value of information with the most cost effective and flexible IT needed to provide it. An approach to information management that seeks to take into account information's changing value, ILM is a process for managing information through its lifecycle.

This report summarizes findings from a year-long study undertaken to explore the views of senior technology managers in addressing ILM. Executive and professional surveys in fifteen case study organizations, and an online survey of 345 IT and storage managers were completed. These interviews sought to understand how technology executives viewed ILM, its business drivers, its implementation record to date, and its ultimate business value. Executives saw the importance of ILM in the drivers and issues underlying it, in the process of aligning business objectives with improved enterprise data management (including policies for data classification and ownership), and in more effectively managing the cost of data stored.

We report on lessons learned in ILM project implementation, and conclude with thoughts on where future ILM research is headed.

II. The Information Problem: Leveraging Growth, Maximizing Value

Paradoxically, in the age of information it is difficult to define information and measure its value. As information technologies move forward, the problem may be getting worse. If we look strictly at the growth of digital information, estimates indicate the total stock of information is growing at over 50% a year. By 2010, assuming the growth rate is stable, the total stock of world information will increase by a factor of six. If the growth rate increases, however, the total information stock will be even greater.

Enterprise data, measured in terms of data volume, is estimated to be growing at 45% a year. This dramatic growth is driven by the enterprises themselves, by environmental factors including ever-expanding government regulation (SOX, HIPAA, etc.) and legal "eDiscovery," and by the steady advance of increasing IT, automation and machine-to-machine communications in core business processes including Enterprise Resource Planning (ERP), Supply Chain Management (SCM), and Customer Relationship Management (CRM) systems among others. Enterprises are by necessity becoming voracious collectors, processors and archivists of information. By one estimate, Wal Mart, whose data warehouse already exceeds 500 terabytes or half a petabyte, has increased its data storage capacity by an average of 32% each year since 1999.[1] And perhaps even more illustrative, Wal Mart's growth rate will only increase as the company pursues data-intensive RFID, "sensor-net" and other supply chain initiatives based around magnifying information intensity and velocity across the chain. A final marker of the magnitude of explosive information growth is the rate of storage spending in user firms despite Moore's Law and innovation-driven declines in per gigabyte storage costs of over 40% a year.

The implications of these dramatic growth rates for companies and for Chief Information Officers (CIOs) are considerable, both above the radar screen and below it. Above the radar screen, the scope of Enterprise Information Management is extensive and expanding, covering many different types of information, information policies, and systems, database and storage management. Across a typical firm, overlapping data repositories are accessed by multiple applications. Whereas multiple OLTP (online transaction processing) systems usually function off proprietary systems and database environments, MIS systems operate across an ODS (online database system) or Data Warehouse serving an enterprise business unit. Even without the inclusion of unstructured or semi-structured data, the challenge is that access, processing, "near line" storage and archiving of information is becoming more complex - information demands from existing data assets continue to grow, while at the same time companies increase

[1] Tallon, Paul, "The Viability of Indefinite Storage: Insights from Email Archiving," Glasshouse Technologies Whitepaper, 2006.

their number of systems, the connectivity between these systems, the data volumes they hold and their requirements for real-time information.

Below the radar screen, solving these issues will require technology, but driven from a business perspective. Business users have always required "the right information at the right time." The question is whether IT is providing it, given:

- We continue to add systems and connectivity between systems at increasing rates
- Data quality issues persist, costing companies heavily in lost revenue and opportunity
- Data security continues to be a major organizational challenge
- Customer insight is key to competitive opportunity
- Users want the "right" information now
- Government regulations continue to expand, and there are serious consequences for executives at all levels if there are data problems.

The paradox facing enterprises, business executives and CIOs today is that information and information management problems have arguably worsened through cycles of technology evolution in IT systems, processing and storage platforms. We are collecting, processing and storing information at unprecedented rates. Where is the value in all of this information growth and how do we extract and use it effectively?

III. The Drivers and Promise of Information Lifecycle Management

It is axiomatic to state that future IT capabilities will be different and driven by better and more effective use of information in the firm. Information Lifecycle Management (ILM) is a policy-based IT systems strategy conceived by the data storage industry to address the problem of enterprise data mobility and storage based on the value of the data. The promise of ILM is seamless information access and storage, where storage efficiency and cost effectiveness are driven by the value of the information stored.

ILM is an ambitious concept, one requiring important changes in management, data policies, and hardware and software technologies to be effective. Our study of ILM focused on three key aspects: first, to understand what is involved in taking a lifecycle approach to managing information in the firm (i.e., what does it mean, what is its utility, what are the costs and benefits of this approach); second, to go into the field to ask senior business and technology managers how they view ILM, its scope, utility, business value, and risks; and third, based on a sample of fifteen companies with varying levels of ILM engagement, to distill practical learnings and early assessments of practices in implementing ILM. The report's research methods, field data collection and company sites, and professional questionnaire are included in the Appendix.

Defining ILM

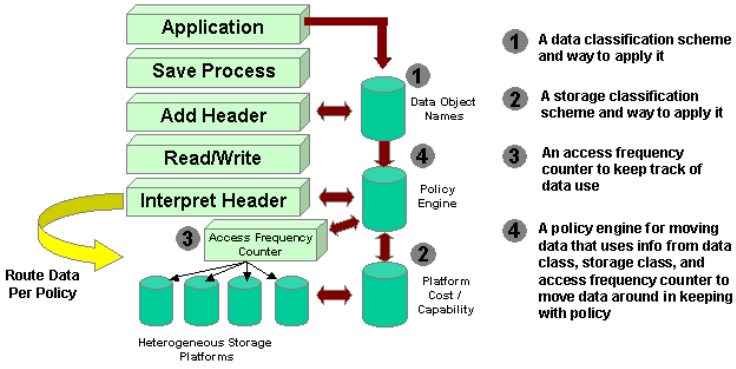
According to the Storage Networking Industry Association's (SNIA) Data Management Forum, Information Lifecycle Management is an end-to-end concept, comprised of the practices, policies, processes and tools used to align the business value of information with the most cost effective and flexible IT infrastructure needed to provide it.[2] Conceptually ILM can be thought of as "a process for managing information through its lifecycle, from conception until disposal, in a manner that optimizes storage and access at the lowest cost." Operationally, ILM can be defined as the application layer assigning value to data, and the data management layer assigning data to different storage resources according to the data's access and protection requirements.

Figure One illustrates a highly simplified architectural view of ILM. It includes a data classification scheme and a way to apply it, a storage classification scheme and a way to apply it, a data access frequency counter to keep track of data use, and a policy engine for moving data around the enterprise using information from data class, storage class, and the access frequency counter to move data consistent with policy.

[2] Storage Networking Industry Association, Data Management Forum <http://www.snia-dmf.org/>

Figure One – Architectural View of ILM

- ILM requires a
  - "front end" to categorize information properly
  - policy engine to request the proper storage category for each information class
  - "back end" to manage the storage infrastructure

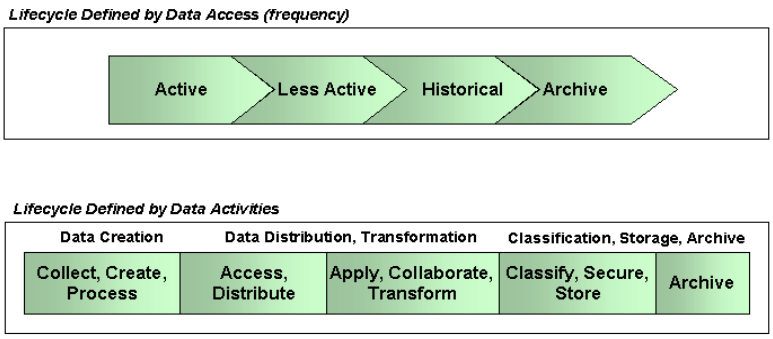


Defining the Information Lifecycle

A generic lifecycle definition includes stages or discrete states of a process or objective, and describes these stages over time. For example, the system or asset life cycle is "the useful or total productive time span of an asset or system. Or the present value total cost for acquisition and operation over the useful life of an asset or system." In terms of how information is accessed, transformed, moved around and stored in the firm, analysis of data movement shows that as data ages, its frequency of access falls sharply. Data access tends to occur within a few days of data creation; after one or two weeks the data is rarely accessed; after 90 days the data is almost never accessed.[3] Thus a simplified data lifecycle can be described as Active, Less Active, Historical, and Archive. Alternatively, the information lifecycle can be described as categories of file and data activities: creation, file distribution and transformation, file classification, storage and archiving. Figure Two illustrates lifecycle definitions by data access and by data activity.

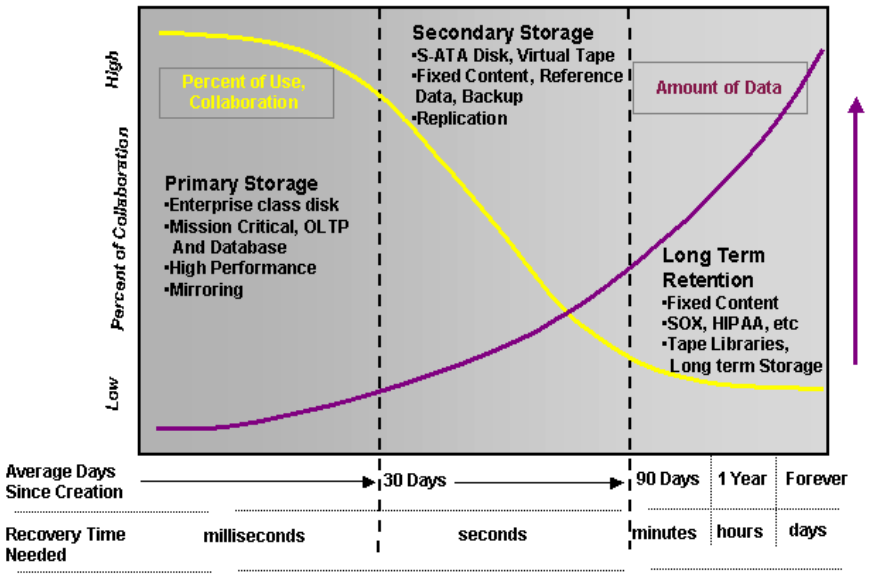
[3] Caveat: not all information falls into this usage profile. An example is firms implementing long tail virtual product shelves, such as Amazon and NetFlix. Here product information is retained for very long periods, although not necessarily on Tier One storage (the most expensive).

Figure Two - The Information Lifecycle



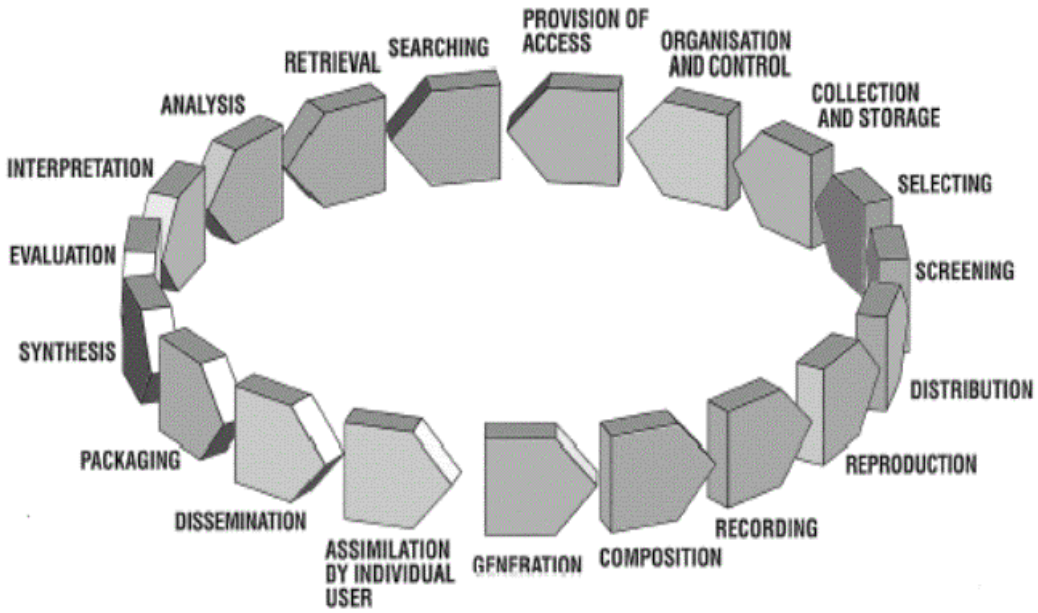
The information lifecycle definitions in Figure Two can be compared with the classic data lifecycle chart pictured in Figure Three. The data lifecycle chart shows a declining data access curve as data age increases, an increase in total enterprise data volume over time, and storage resources allocated to different data access values (frequently accessed, less frequently accessed, infrequently accessed historical and archive data). This chart best describes structured data (transactional data). It is problematic to apply to unstructured data for reasons we will address later in this report.

Figure Three - The Data Lifecycle



A final model highlights information value-adding processes. This model was originally based on the analysis of activities of information enterprises (advertising, consulting and information services companies for example). Figure Four shows a cycle that begins with the generation of ideas and the activities adding value to information including collection and storage, organization, information search and retrieval. It includes processes for analysis, interpretation, evaluation, synthesis, repackaging and dissemination across the lifecycle[4]. None of these models is put forward as ideal; each represents a part of how information can be defined by how it is used or how it is created, added to, and distributed across the enterprise.

Figure Four: Information Value Processes



The Factors Driving ILM

What are the main business drivers of ILM? Respondents in our field study identified multiple drivers, including exploding information growth in all data formats, legal discovery and regulatory compliance, Web 2.0 collaboration, search across the enterprise (including stored data), content generated by enterprise applications, paper and electronic

[4] Browne, Mairéad, "The field of information policy: 1. Fundamental concepts," Journal of Information Science (1997, Vol. 23, p. 261). http://jis.sagepub.com/cgi/content/abstract/23/4/261

content integration (example: bank credit files), and access and storage from more devices. Underlying information trends include the move to greater real-time information, consolidation of data centers and IT infrastructure, a focus on middleware and lifecycle management of information and applications.

As reported by our respondents, a summary of the main drivers of ILM include:

- *Enterprise Data Growth*: We earlier noted that enterprise data is estimated to be growing at 45% a year (in terms of data volume). When asked about storage growth, respondents noted the growth of systems and applications are driving demand for more storage. Audit and compliance are major drivers for archival storage.

- *Growth In Unstructured Data*: The growth in enterprise data is primarily in unstructured data. A soft estimate is that 80% of all enterprise data is unstructured.[5] IDC recently estimated that over a quarter of all enterprise data lies outside the corporate and business unit data centers. One of the main problems with unstructured information is that its value is defined by the user, not the enterprise (or IT system).

- *Relational DBMS Performance*. Several companies reported that information growth is taxing the practical limits of relational database, data warehouse and network scalability and performance. While processing power doubles and disk prices half every 18 months, no such scalability model applies to data management software.

- *Information Access and Security*: Data growth is also taxing the volume of information that firms can effectively use and secure. Many manufacturing firms do not even try to collect the information their factory floor equipment can generate. And we are at the beginning of sensor and other forms of Radio Frequency Identifier Devices (RFIDs) deployment into the chains, for example, for purposes of managing the flow of goods and information in real time. As real-time systems depending on more types of information (manual and machine based) are deployed into supply chains and other integrated business processes, data use, access and security demands increase the need for continual improvement in storage software and management control.

- *Data Classification*: IDC estimates that less than 10% of all enterprise information is classified or ranked according to value. The company also estimates that the amount of classified data will grow by 50% a year. But even if the percentage of classified data is growing, the growth of all information in the enterprise is growing much faster. Therefore the percentage of classified data as a percentage of the growth of

[5] I refer to the 80% unstructured data figure as soft as I have been unable to cross-check its accuracy. It appears this number has been used in IDC reports but no explanation of how it was arrived at was provided.

unstructured information is less than 1. Unless data classification methods improve dramatically and are deployed effectively, the ratio will worsen.

- *Maximizing Productivity*: As firms continue to add real-time systems, applications and databases, the fixed costs (Total Cost of Ownership or TCO) of enterprise storage systems continue to escalate. Viewed from a storage systems perspective, the impact of this investment on firm productivity is not well known. Research is underway, but it is early days in modeling the productivity impacts of IT systems investments.[6]

[6] Aral, S., Brynjolfsson, E., and Wu, D.J., "Which Came First, IT or Productivity? The Virtuous Cycle of Investment and Use in Enterprise Systems," 27th Annual ICIS Conference, Milwaukee, Wisconsin 2006.

IV. Implementing ILM: What Companies Are Doing, What Managers Are Saying

Of the fifteen companies participating in the study, five had active ILM, data warehousing and business intelligence initiatives underway. Four had started the planning steps required for an ILM, data warehousing, or business intelligence effort (which included ILM). Three companies were responding to specific legal "eDiscovery" directives. And all firms in the sample had some form of compliance, data retention or security projects underway, ranging from major initiatives in the financial services and healthcare firms, to smaller efforts focused at the level of business units or departmental functions. Typical of the smaller efforts were email archiving and customer records management, where some form of ILM "extender" products were being evaluated or deployed.

As expected, the organization of ILM projects varied widely according to the goals, level and sponsorship of the effort. As a general rule, multi-function (enterprise level) ILM and data warehousing or business intelligence (BI) efforts required relatively large sponsorship networks and executive level commitments of time and stewardship across the full planning and project lifecycle. In the words of one executive, "these projects can be significant and require the active support of top management… they (generally) involve customer data and the information necessary to meet functional or personal goals, or they arise out of legal or compliance directives that usually require immediate attention… you don't want to be starting from scratch on one of these."

No dedicated Records Information Management (RIM) projects were uncovered in our case sample. However, all firms had some activity underway in data security or compliance, with records management an important project activity. Electronic records pose a business risk associated with potential exposures in legal discovery, regulatory inspections, industry investigations and privacy rule violations. However, records

management still appears a downstream activity, functioning as an added management layer in organizing data and information for longer-term storage and data archiving.

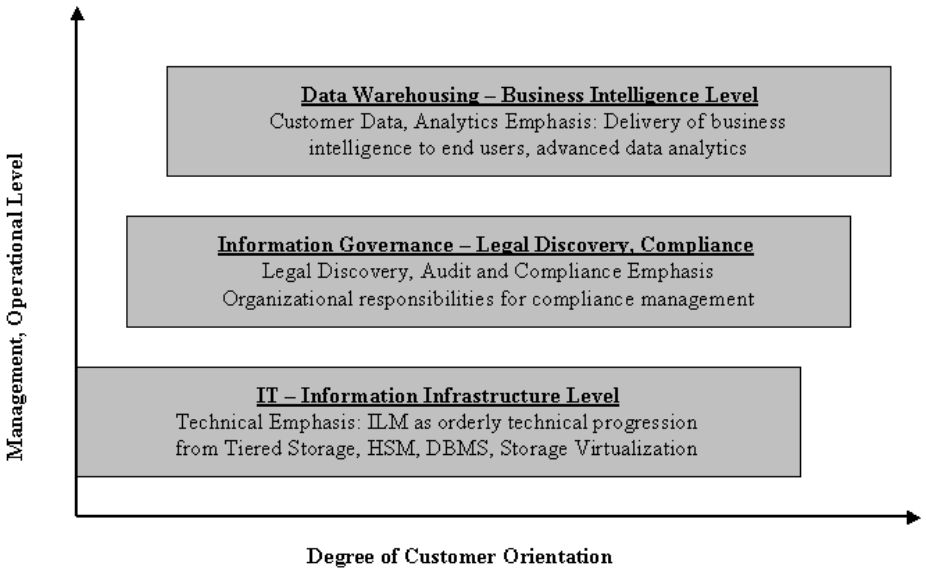
Three principal business value objectives were identified in interviews:

- supporting the business in improving performance, especially in customer facing activities and improving customer relationships and interaction
- improving performance of IT through better alignment of data and information services with business needs, including lowering costs, improving quality, and improving the performance of information access, search and storage
- strengthening the foundation (people, systems, records) for ensuring compliance ("decreasing the business risks and costs of compliance").

Other important objectives included working with the business to achieve a more cohesive ("disciplined") approach to the use and storage of information in the firm. Technology and IT infrastructure objectives included reducing IT complexity in database, decision support and storage systems; implementing improved business intelligence applications (improved query and search capabilities, improved response), and improving cost performance and decreasing the TCO (total cost of ownership) of storage systems and archiving as a percentage of total IT spend.

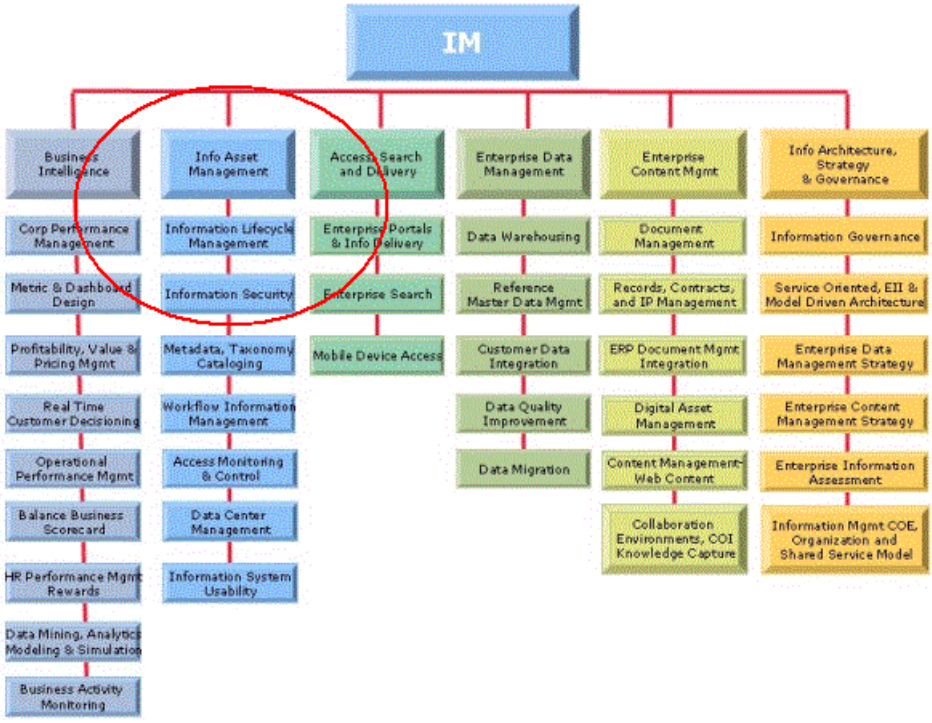
Figure Five summarizes the main categories of enterprise IT initiatives where some form of ILM project activity was underway.

Figure Five: ILM In Enterprise IT



A final model of how ILM fits within the practice of information management (IM) is illustrated in Figure Six. The figure defines ILM as a practice area within Information Asset Management, complementary to other IM practice areas including security, data classification and workflow management. Again, this model is not being put forward as ideal; it is one representation of how ILM can be defined within the broader IM practice area.

Figure Six: ILM In IM Practice



What Companies Are Doing: Example ILM Projects

We observed different aspects of ILM embedded in the larger-scale IT initiatives reported above. At the project level, we did not find a single or even a few approaches to how companies were evaluating or implementing ILM. Rather, the view was that ILM was a functionality or project activity as part of specific initiatives in compliance, data retention, or legal discovery, or in customer data and data warehousing projects, or in more technical storage projects ranging from email archiving to long-term digital records archiving in, for example, healthcare or financial services companies.

Below is a summary of project activities with comments from survey respondents:

Regulatory Compliance, Records Management, Deep Archiving

- "Legal advice (discovery) retention policy - especially electronic forms - i.e. email. When to archive, and when to destroy"
- "Architecting a Sarbanes-Oxley (SOX) workflow system complete with storage management components"
- "SOX related first attempt at DRM planning and records management"
- "Project addressing costs and risks of deep archived data, including security, access (search), and cost for legal discovery."

Customer Records Management, Data Warehousing

- "A CRM tool to capture, input and centralize all disparate client information. This project will allow us to more efficiently manage our resources and better serve our client base"
- "ILM within the context of CRM (customer relationship management) and contract lifecycle management"
- "Revamping the workflow of customer data using a lifecycle approach (applications and information)"
- "Looking into how and where ILM fits into our data warehousing and business intelligence projects. No formal project but an area of IT management interest."

Data Classification, Data Migration

- "Developing a data classification system to further develop and improve policies related to protecting and storing data"
- "HSM (hierarchical storage management) project that allowed for data movement from the most expensive disk storage platforms to expire and be moved off to less expensive forms of data storage"
- "Purchased new backup and restore software that has a data migration piece to it, and we are trying it out in the IT department for approval for corporate implementation"
- "Architect of large project for automatic movement of data based on the average access of data at the block level"
- "We are beginning the process of classifying our data, or, more properly stated, working with the data owners to classify it. After classification is done we will be managing it based on what we discover."

Email, Application Archiving, Data Repositories

- "We have an Enterprise Vault project underway to migrate email and files to lower cost disk and tape systems."
- "We are setting up an enterprise repository for documents, reports, images and files using a form of information life cycle management."
- "IT is lead in implementing ILM in a large document repository where one of the goals is to store less accessed data on less disk."
- "We are in the middle of a project to pick an enterprise records management solution mainly focusing on electronic data (file systems, database, email).

Data Center, Storage Systems Consolidation, Backup and Restore

- "We have a large project underway looking into consolidation of all storage use onto a common storage platform / system. The goal of the project is to develop a common backup/restore policy and a business continuity approach to all storage."
- "Consolidation of file servers to Network Attached Storage (NAS). We do not want to take years worth of junk with it."
- "We are currently re-architecting our BURA (Backup, Restore and Archiving) system in order to improve data availability and storage efficiency. ILM is being looked into as a way to help set policies for data recovery Service Level Agreements (SLAs)."

What Managers Are Saying: Results from Professional Surveys

In addition to detailed interviews with executives in our fifteen field case sites, we also completed an online, professional survey of 345 IT and storage professionals, ranging in titles from Chief Information Officer (CIO) to Systems and Storage Managers in large user companies. Respondents were asked to complete a half-hour, online survey through their membership in the Storage Networking Industry Association (SNIA), the American Records Management Association (ARMA), through membership in StorageNetworking.org (the Information Storage Industry Center's (ISIC) community of practice), or through ISIC sponsorship directly. Survey methods and procedures are outlined in the Appendix. Approximately 3,500 individuals were notified about the survey, 376 completed surveys were received, of which 345 were suitable for analysis.

The majority of survey respondents defined ILM as a policy-based approach to improving records and information management in their companies (52% N=181). Slightly less than half of respondents, however, saw ILM as a technical and systems management issue, following innovation-driven improvements in storage systems and software available from the industry (40% N=140). When respondents were asked to rate

the importance of topics often associated with ILM in trade and industry sources, their responses clustered into two groups. As illustrated in Figure Seven, respondents rated Categories 1-3 (Tiered Storage, Storage Resource Management, and Data Archiving) highly. Similarly, Categories 8-10 (Data Protection, Records Management and Compliance) were rated at the top of the one to ten point scale of importance.

Figure Seven - ILM Topic Importance

"Following is a list of IT topics that are often associated with ILM projects found in industry.
Please rate each topic in terms of how important it is in influencing your own company's approach to ILM.
10 is the highest in importance, and 1 is the lowest."

Importance Rating	Lowest →							Highest		
	1	2	3	4	5	6	7	8	9	10
Tiered Storage, HSM	3.7% 13	3.2% 11	4.0% 14	2.6% 9	12.1% 42	10.3% 36	14.4% 50	19.5% 68	16.1% 56	12/6% 44
Storage Resource Management	2.0% 7	2.3% 8	3.2% 11	6.3% 22	15.2% 53	12.4% 43	17.2% 60	18.4% 64	13.8% 48	8.0% 28
Data Archiving	1.7% 6	1.4% 5	2.3% 8	4.6% 16	7.8% 27	11.8% 41	14.4% 50	23.3% 81	16.4% 57	15.2% 53
Application Archiving	6.9% 24	6.6% 23	12.4% 43	9.5% 33	19.0% 66	12.1% 42	13.2% 46	9.8% 34	5.2% 18	4.3% 15
DB, Application Integration	2.9% 10	5.5% 19	5.2% 18	6.0% 21	15.2% 53	11.8% 41	15.5% 54	17.2% 60	10.9% 38	8.0% 28
DB Search, Data Mining	3.2% 11	5.5% 19	7.2% 25	8.0% 28	11.5% 40	12.4% 43	17.5% 61	15.2% 53	10.6% 37	7.5% 26
Data Warehousing	4.0% 14	6.0% 21	8.6% 30	9.8% 34	13.8% 48	10.6% 37	15.8% 55	12.9% 45	10.1% 35	7.2% 25
Data Protection, BusinessContinuity	2.3% 8	0.3% 1	1.7% 6	2.9% 10	6.6% 23	4.9% 17	7.5% 26	20.7% 72	21.8% 76	29.9% 104
Records Management	3.2% 11	2.6% 9	3.7% 13	6.9% 24	12.6% 44	11.8% 41	17.0% 59	17.5% 61	12.4% 43	11.2% 39
Audit & Compliance	4.3% 15	2.9% 10	3.2% 11	6.6% 23	8.6% 30	10.1% 35	10.9% 38	16.4% 57	16.1% 56	19.3% 67

The pattern emerging from these responses was that managers were of two minds – technically focused respondents were split on whether ILM was technical or policy; management and policy focused respondents tended to view ILM as policy. We interpret this finding to mean that with the relative newness of ILM and inexperience in field implementation, respondents anchored their views of ILM in areas they know well.

The Benefits and Drawbacks of ILM

We also asked respondents what they saw as the main benefits and drawbacks of ILM. Most saw the main benefits of ILM as increased control over data, as insuring regulatory compliance (minimizing business risk), and as potentially reducing costs (by eliminating redundancies in data storage):

- "Knowledge and control of valuable, critical company data assets"
- "Policy based storage"
- "Regulatory concerns limiting liability, compliance with data retention regulations"
- "Controlling storage growth"
- "Ability to manage information in a business value chain context"
- "Opportunity to create consistent, repeatable, efficient business processes for data management"
- "Long term cost savings."

Conversely, some respondents felt there would be no immediate benefit to ILM. Rather, they noted the importance of other organization and systems work that needs to precede ILM before ILM can have any tangible positive impact:

- "The benefits of ILM are not yet fully available, meaning that adequate organizational and categorization tools that can handle the varying rules of retention and appropriate disposal efficiently… are not in place."

The picture that emerges is that managers see the primary near term benefit of ILM as improved management control over data and better storage purchasing decisions:

- "A sense of order and the ability to 'understand' our data so that we don't over-purchase hardware or follow the 'tack on more' methodology when we think we're short of storage space."

However, approximately half of all respondents felt that any benefits accruing from ILM would have to follow important systems and organizational changes beforehand.

ILM Drawbacks

While respondents cited cost savings, improved management control of data, and improved compliance as potential benefits, others saw the same list as potential problem areas: increases in cost and complexity (of the storage management environment), lack of standards and confusion in the marketplace, and required up-front investments in data, applications and storage hardware:

- "Very large scope. It means too many things to different people"
- "Upfront analysis of processes, upfront costs of implementing standard processes"
- "Costs, lack of management understanding of the risks and rewards"
- "Cost and the fact that there is no real standard for archiving"
- "Difficult to manage. No complete solutions, many vendors attacking pieces"
- "Lack of standards to protect buyer investment."

Respondents also cited potential problem areas familiar to all large-scale IT systems development and infrastructure projects:

- "Difficult to implement. Categorization of data is labor intensive and complicated."
- "1. Time to ramp-up competency 2. Lack of standards 3. Ability to reap benefits along the way."
- "Pushback from the workforce (employees) and the unwillingness to change."
- "Too many solutions tie in hardware to what is essentially a soft problem."
- "Transparency – ILM crosses multiple business units. This makes it a harder sell to management."
- "Today ILM presents a full additional layer of infrastructure and management. Until a company reaches a very high threshold of usage, its value is difficult to quantify and communicate to management."

The picture emerging from survey responses is that managers see as many problems with ILM as potential benefits. And while proponents base their arguments around the conceptual soundness of viewing information as a corporate asset, and in aligning business demands with real-time information and the IT infrastructure needed to provide it, ILM is seen as an ambitious concept whose value needs to be proven in effective implementation. As one executive stated, "it is important to start somewhere, whether it is looking back at your infrastructure, or looking out to your customers and the support for the business needed there… Whether it is called ILM, or something else, is less important than working with the business to get it done."

V. Management Issues Posed By ILM

Whether labeled "information lifecycle management" or "data management," the problem of escalating information demands and long-term storage and archiving of information is upon us. For better or worse, IT has the primary, if not sole responsibility for managing, protecting, storing and leveraging information across the enterprise. For many IT organizations, an effective ILM strategy begins with a thorough assessment of the business environment and the IT storage environment and ultimately hinges on the effective communication between the two, something in the words of one executive interviewed, "they will have to get used to."

Certainly there are many ways in which an enterprise can attack the problem of exploding information and storage growth. Like many concepts trumpeted in the past, such as enterprise data modeling and data engineering, ILM implementation will require difficult cultural, organizational and development changes. These changes can include:

- Central Management — ILM may require the IT organization to reorient or restructure itself to facilitate the new focus. This can mean centralized IT management groups with an end-to-end focus on data management versus infrastructure management.

- Heterogeneous Scope — An enterprise ILM strategy must adapt itself to the reality that data exists throughout the entire organization on a variety of computing, network and storage platforms.

- Data Value Alignment — Managers must vary the approach for managing different types of information. Across the organization, policies must be created to define the value of data, what retention rules apply to what data, and who can access what data. Then, the storage infrastructure must be aligned to reflect the appropriate value, retention and access policies for classified data.

- Metadata Creation — A required behavior change for most information creators and users will be the provision of metadata, or descriptive information about the character, function and value of content. Metadata specifies who created the content, for what purpose, and what attributes of the content dictate how it should be managed over time. In current practice ILM is being carried out in a rudimentary way by employing generic rules to move, archive or delete content, rather than by content evaluation. Examples are the company that deletes email after 30 days and the financial services company that archives all email attachments that have not been read for 90 days (regardless of the content of the email or attachment).

- Financial Incentives: Paying For Value — Of course, even if users can be convinced to create metadata or to help in verifying its automatic creation, there is nothing to prevent users from thinking all of their content is mission critical forever, and therefore should be maintained on the highest performance (highest cost) storage device. Chargeback systems of varying type and degree have been introduced over the years to give economic incentives to users to optimize their use of IT resources.

Choosing the Starting Point

If ILM can be distilled down to the need to classify data according to its value, define and set strategic data policies, and provide accurate cost modeling to implement effective service level agreements (SLAs), then most companies should:

- Create tiered storage hardware to enable different services and service levels, offering potentially different cost structures.

- Categorize data (e.g., by application) to describe its specific value to the business process or organizational unit. This process matches data classifications with their proper tier in the storage architecture.

- Layer storage management software onto the environment to enable reporting, protection, and data migration capabilities.

These capabilities provide the starting point for most ILM implementations. Once underway, ILM initiatives typically follow a phased approach:

- Implement tiered storage architecture for ILM. This facilitates placing information and applications in the appropriate location for storage management automation.

- Implement application-specific ILM. Most companies then focus on a specific application, often one requiring large-scale IT resources (for example, an ERP application) or with the potential for rapid ROI (for example, a CRM application).

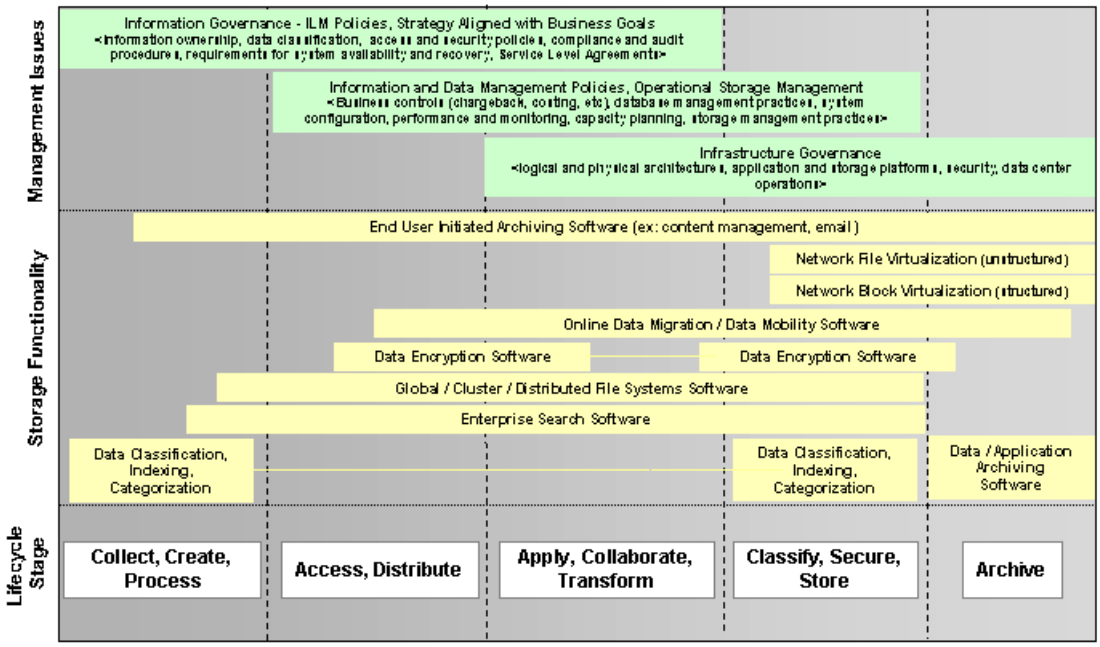
- Implement cross-application ILM. Companies then focus on extending ILM across multiple applications.

Success measures typically involve both cost savings and improvements in availability, recovery and performance of the application and/or storage system.

Mapping Management Issues by Storage Resources by Lifecycle Phase

A question of interest throughout our study has been to ask what management issues arise at different stages of the information lifecycle? We asked this question of respondents in both our case sample and in our professional surveys. The analysis of their responses is illustrated in Figure Eight. The figure contrasts top level management issues (as seen by ILM project teams), storage functionality (in terms of storage applications and IT functionality), and a functional definition of the information lifecycle.

Figure Eight: Management Issues by Storage Functionality By Lifecycle Stage



ILM and Information Policies. The top level management issues are alignment of ILM objectives, strategies and business goals, information management policies defining operational storage resources and storage management, policies for information classification and across enterprise data migration, and infrastructure management, which includes application and storage platforms, security, and data center operations.

Storage Resources and Applications. The middle level of the figure groups the main storage capabilities – the software and hardware that make up the tiered storage environment. This includes capabilities and applications such as data classification,

search, data mobility, data encryption, network file and network block virtualization, and end user archiving software. This list was developed by analyzing project initiatives and interviews with storage administrators and ILM project teams, along with compilation and analysis of secondary information published on ILM.

Lifecycle Stage. The information lifecycle depicted here is simplified. The linear block sequence of activities captures most, but not all, of the sequence of activities in discrete stages of the information lifecycle. Stages are not necessarily sequential.

An important qualifier to this mapping of the information lifecycle is that it does not address differences between unique and replicated information in terms of growth, value, lifecycle, or management. Most conceptual discussions of ILM side-step the issue of unique versus replicated information in the enterprise. In an ideal world, we would most likely be interested in the original creation of a piece of content, and in its final form, in that knowledge of both "starting point" and "ending point" allows classification, identification of the activities over time adding value, and description of the final product for storage and archiving. Intermediate drafts, replicates, etc., can be safely discarded unless there is some value in a record of the value-adding process. We lack ways to separate out and measure unique information production and growth from replicated information growth. However, as our ways of thinking about information and our abilities to track, measure and manage information in the enterprise improve, this will be an active area for future research.

Business IT Alignment Issues in ILM Projects

We now move from general management issues to ILM project issues reported in our case studies and professional surveys. Responses were analyzed and pooled into five issues of alignment between business policy and storage and IT practices:

- *Aligning information value with storage costs*. ILM is concerned with aligning information value with the cost of storage. The primary motivator to migrate data from high speed – high availability storage devices to slower and less available storage is cost. And while most IT organizations have analyzed the cost differentials between disk, tape and other storage media for different classes of data, ILM requires better and more extensive cost modeling of data migration across storage resources.

- *Aligning storage availability with user activity and demand*. Data activity and data reuse decline, on average, over time. However it is not a smooth decline. There are spikes in the access and use of data over its lifecycle. Ensuring that users throughout the enterprise have access to the information they need depends on some measure of user and use classification.

- *Aligning records management and data retention and disposal with new regulations.* Records management is comprised of both the retention period and archiving (read only) of enterprise information. Both are defined by regulation. For example, in discovery firms must demonstrate that "best efforts" (good faith efforts) have been made to insure compliance with regulations such as SEC Rule 17a-4, which mandates email retention. There are situations, however, when data reduction is more appropriate. Privacy legislation, for example, may require a reduction in data retention time in order to provide individuals a higher degree of data protection. In practical terms, evolving regulation has many instances of conflicting rules for data retention and data storage. Civil case law does not provide clarification since it is still in its infancy. Yet companies are legally required to formulate data retention and storage / archiving policies which resolve in policy terms regulatory inconsistencies.

- *Aligning search and data recovery with data value.* Data recovery relies on the prioritization used to define the distribution of data across the most efficient storage resource configuration (availability, access, recovery time, cost). A primary factor guiding data availability and retention policies is business continuity. Business continuity is concerned with vertical (example: CEO versus department managers) and horizontal (example: finance versus sales) data value. Data recovery time is a function of the business need ("value") of that data. In certain circumstances, business continuity is also a function of compliance requirements, such as Sarbanes- Oxley (SOX) or HIPAA. An off-shoot of ILM has been lateral pressure on backup and restore and archiving software companies to incorporate greater data classification and search functionality in their software.

- *Aligning access controls (security) with user need and data value.* Assigning data access and security levels for individuals in the enterprise requires assigning the type of information (and application) that the user requires, again reinforcing the need for some measure of user and use classification of data.

The above list of alignment issues emphasizes a storage centric view of ILM. This approach may have tactical advantages for many enterprises in the early stages of ILM. However, according to our interviews, a storage centric view is also seen as limited and needing expansion through a broader information integration strategy. Nonetheless, knowledge management, content management, digital rights management, and the long list of other technologies needed to extract greater business value from enterprise data are still seen in varying stages of maturity.

VI. Summing It All Up: Twelve Lessons from ILM Practice

The Imperative

The rate of information growth in the enterprise and the firm's business network of customers and suppliers is voluminous and increasing. Industry reports, university studies, corporate white papers and industry conferences all voice an escalating challenge for business organizations: how to address and manage information growth; how to find, extract and use the value in information; and how to develop legally responsible and cost efficient systems to store valued information and the rest. By their own admission, companies cannot keep adding management and IT resources indefinitely to store all of the information and data collected and produced by an enterprise. Worse, there is little security in believing that as enterprises allocate more resources to store greater amounts of information the value of stored information increases. To the contrary, business and IT executives interviewed in this study were worried that under current incentives and policies, costs, business risk and the potential for greater risk was increasing.

If respondents are painting an accurate portrayal of increasing risk, enterprises may be well advised to take a step back and think about the problems inherent in dramatic information growth in new terms. Where is the value in capabilities to classify, store and deep archive a constantly growing percentage of information in the enterprise? Should there be a rule that for every project focused on retaining and adding information stores to the enterprise, there is a parallel project on discarding unused information?

The discussion points and questions addressed in this report surfaced in many interviews with business and IT management over the course of the research project. In interpreting and analyzing lessons learned in early ILM initiatives, twelve emerged. They are:

Lesson One: Focus on Information Value and the Processes Used to Extract Value

ILM has been useful in focusing management attention on the current state of innovation in Enterprise Information Management. It has helped illuminate the diversity in business forces involved, ranging from the legal requirements to keep information, to the risks (and costs) of keeping that information indefinitely without effective or widely distributed policies for valuing that information. Technological factors have made it easier and more efficient in the short term to use technical means to store more and more information. This is very much a two-edged sword. Management factors, including business risk, compliance, and business continuity of stored information now make it desirable for business and IT to build new partnerships to address the issues of valuing information, from its collection and use to storage, to discarding information.

Lesson Two: Focus as Much on Discarding Information as Storing It

ILM has also helped bring into focus the problem of discarding information in the enterprise. Moore's law and business pressures for more information and more and better analytic processing of that information (data warehouses, data mining, business intelligence, business analytics) had relegated the problem of discarding information to the IT background. Security, theft, compliance and the business risk of archived information has moved policies for discarding data back into the foreground. By their own admission, the majority of companies in our field sample had inconsistent (and incompatible) policies for discarding information. Moreover, there were questions regarding the effectiveness of policies to discard data, given the practical realities of data replication, multiple copies, backup and recovery copies, archival copies, and so forth..

Lesson Three: Initiate ILM Through Policy and People, Not Technology

ILM is a management process implemented through policies, by people, supported with technology, in that order. Companies in our field sample were skeptical of the value of technology initiatives proceeding under an ILM banner without clear business sponsorship and active line manager involvement and support. However, absent of clearly defined ILM project benefits, IT and storage professionals cited many problems in gaining commitment from the business to initiate ILM activities. This is the most likely reason that ILM projects often fall under the umbrella of larger IT initiatives with established business sponsorship.

Lesson Four: Recognize that the Primary Drivers of ILM Are Compliance, Legal Discovery, Risk Management and Data Retention

In our field sample and professional surveys, the main business drivers for initiating ILM projects were compliance, legal discovery, risk assessments for archived information, and security and regulatory requirements for data retention. The magnitude of these pressures has resulted in a proliferation of tactical responses - the projects cited earlier in this report identified many of these - and a focus on one application, email, as the application to start with.

Lesson Five: Acknowledge that Email's Value is User Defined, Not Enterprise Defined

Email presents the difficult problem that value is defined in the user's terms, not the enterprise's. And user studies of email show over and over again that users are very reluctant to discard old emails. Why is this? For many users, email functions as a personal work record and proxy file storage locator. It is not only a means of communication but a means for record keeping and storage of information (context,

content and use). Key data files are often stored as email attachments and email threads are often a source of user and organizational memory. Charge-back systems are used to allocate enterprise storage resources. What would be the implications of charge-back being used to price and allocate storage resources at the user level? For example, applied to inbox capacity? If users start purging email records, would this be the desired outcome at the enterprise level? What are the risks in taking this approach (inappropriate email deletion)?

A 2006 InformationWeek survey concluded that over 65% of firms interviewed had not adopted a formal email policy. Users had little sense for what emails needed to be retained. Yet in terms of the business, these emails constitute a record of past events. And in the absence of a formal policy, there would be no way in which appreciation for the financial and reputation risk firms face from inappropriate email handling could be communicated throughout the firm. Examples from Merck's battle over the adverse side effects of Vioxx, or Morgan Stanley's alleged fraud case involving Sunbeam highlight the necessity to review email retention and formal retention policies.

Lesson Six: Recognize that Application-Specific ILM Implementations Can Disrupt Enterprise-Wide Initiatives

The vision of ILM may be enterprise-wide and holistic, but firms are deploying application-specific solutions in response to short-term needs, such as complying with data retention regulations, or improving storage utilization. Given the complexity of any scaled-up ILM initiative, this pattern will hold for some time.

The primary operational components of ILM in the storage-centric view (which we have earlier stated is tactically efficient), are tiered storage, data classification, and data movement. Tiered storage is the basis for delivering multiple, differentiated service levels (SLAs), across physical types of storage.

Data classification is the process of sorting data into meaningful groups and applying lifecycle management policies. This requires metadata, or the equivalent information about data that allow it to be described and qualified, either manually (expensive) or automated (imprecise). Once metadata has been defined, software is needed that can apply and leverage it.

Policies need to be shaped by the business. No single group has a complete sense of what can or should be done. The businesses in our field study relied on enhancing communications and a consensus-oriented approach to ensure data classification and data lifecycle policies reflected cross-functional concerns.

Finally, data movement migrates data between storage tiers to accommodate changing service level requirements. This capability is key for dynamic lifecycle management. Ideally, data movement is an automatic and non-disruptive process handled by intelligent software – meaning that human intervention is not required once policies are programmed, and data remains accessible to applications during and after data movement. This capability is easily stated but difficult to implement, and would need to work effectively with other ILM components including data search and retrieval, and flexibility in applying levels of data protection and recovery.

Tactical triggers for initiating ILM projects include:

- Meeting business and regulatory requirements for data retention and access
- Defining and applying appropriate storage service levels
- Improving application and file system performance
- Lowering storage hardware costs
- Speeding up data management operations like backup and restore, upgrades, or replication.

Lesson Seven: Implement Storage Best Practices

Large-scale ILM implementations are years off. However, ILM initiatives usually start with a tiered storage architecture, where mission critical transactional data is stored on high performance disk systems attached to servers (online storage). Less critical data, such as months-old sales or inventory data, might be stored on a storage area network made up of less costly, slower drives (SATA). And document back ups from enterprise PCs (examples: Word documents, spreadsheets) might be stored on archival tape systems (performance-automated, access-centric, or density stacked tape). Figure Nine illustrates the design of a tiered storage architecture, and the performance tradeoffs across different storage tiers.

Figure Nine – Information Storage Architecture, Resource Attributes and Example Service Level Agreements (SLAs)

COST ↑

Storage Tier	Protection	Performance	Availability	Recoverability	Scalability	Application Profile	Archive (restore)
Tier 1 - High Availability (HA) Premium Near Line Storage Disk	Remote, Synchronous, Dual Mirrors [RPO* less than 5 minutes]	Fast I/O 800+/sec, < 2 second response time	Enterprise Class (99.999%)	Cross Site Replication (RTO < 5 minutes)	High Dynamic	Real-Time Mission Critical	Archive Performance Tape
Tier 2 – Enterprise Disk	Mirrored [RPO less than 15 minutes]	Fast I/O 800+/sec, 2 second response time	Enterprise Class (99.99%)	Cross Site Replication (RTO < 15 minutes)	High Dynamic	Mission Critical; Corporate Systems	Archive Performance or Access Centric
Tier 3 – Midrange Disk	RAIDs [RPO less than 12 hours]	Moderate I/O Between 100 and 800 transactions / sec, query	High (99.95%)	RTO < 12 hours	Moderate	Mission Supporting, Corporate	Access Centric
Tier 4 – Serial ATA, Network Attached Storage (NAS)	Alternative Site Recovery [RPO less than 24 hours]	Moderate I/O or No SLA Transaction Rate	Near High (99.9%)	RTO < 24 hours	Low (Fixed)	Corporate, Development	Density Stacked Tape

RPO – Recovery Point Objective
RTO – Recovery Time Objective

Sources:
Adapted from Industry White Papers, the Data Management Forum, Storage Networking Industry Association (SNIA), and field interviews

Lesson Eight: Define Lifecycle Classification and Data Movement Processes

The current state of data classification is largely a byproduct of hierarchical storage management (HSM), where the primary classification criterion was data age. Most data today is classified by access or availability, recovery, and cost. New data classification software regimes address a wider set of data attributes and data uses [See Figure Ten].

Newer classification software is designed according to four principles of data access, data use and migration:

- Access and initial placement define the performance characteristics seen by the user in the application
- Recovery and protection define what happens in the event of a primary data failure
- Discovery, retention and disposal define the service characteristics of archived data
- Security and access control defines how data is protected from unauthorized use.

Figure Ten - Data Classification Criteria

I. Access and data placement

- **Defines the performance characteristics seen by the user in production applications**
  - Where is the data placed initially? How accessible is the data at its point of creation?
  - What are the initial performance requirements?
  - What is the trigger event that can change the status of the data and will result in a change of requirements or a change of classification

II. Recovery and protection

- **Defines what happens when business data is damaged or destroyed**
  - In the event of a failure of the primary data availability, how quickly can the data be recovered?
  - What methods and levels of data protection are required to meet recovery needs?

III. Discovery, retention, disposal

- **Defines the service characteristics of data that has been archived**
  - How quickly can archived data be accessed?
  - How broadly can contextual search be applied?
  - For how long and under what conditions will data be stored?
  - What is the trigger event or timeframe that initiates disposal? How is disposal accomplished? How is the audit trail managed?

IV. Security

- **defines the overall management of the data protecting it from unauthorized use**
  - What access control, physical protection and encryption will be employed?
  - How will this change as the status (value) of the data changes?

Sources: Adapted from Industry Sources, Field Interviews (StorageTek, Best Practices in Data Classification for ILM, January 2005)

Lesson Nine: Lead Technically with Assessments of Search and Automated Classification Software (including metadata classification)

Two of the most important technical and policy aspects of ILM are search and data classification. Improvements in these two technologies are necessary to advance the problem of defining information value, and of how to move and store information according to its changing value, especially more complex multimedia data (text, image, audio, video). The importance of search is in the combination of advanced search with externally supplied metadata. The latter can be created manually or automatically by auto - categorization software. The importance of data classification is it defines the process for assigning information attributes, whereby data can be grouped into logical categories by business objectives, application, and physical (logical) storage system.

Lesson Ten: Review eDiscovery Preparedness

On December 1, 2006, new US Supreme Court rules for electronic discovery of documents in civil cases took effect. The rules specify requirements for submitting electronic documents, including e-mail and other forms of communication as evidence in

civil cases. A ComputerWorld survey of 170 IT managers in December 2006 reported that 42% of IT managers surveyed did not know if their company was prepared; a third said their company was not prepared.

Lesson Eleven: Assess The Business Case For ILM

In the end, ILM is not a software solution but a collection of processes complemented by software and hardware. If ever fully implemented, ILM's promise is to help organizations better manage their data (and therefore the IT infrastructure that manages the data) from the time that data is created until it is no longer needed. On the hardware side, tiered storage enables different services and service levels, offering potentially different cost structures. On the process side, data is categorized (by application) to better describe its specific value to the application and to the business process. This serves to match classifications of data with the proper tier of the storage infrastructure. Finally, there is the software infrastructure layered onto this environment to enable reporting, protection, and data migration capabilities. Only when these are available, in total, can ILM be fully implemented, and its benefits realized. To be effective, ILM requires a strategy for adoption and implementation.

Lesson Twelve: Recognize that there Is No One ILM Blueprint, Although a Direction is Evolving

Our analysis reported here found there is no single, clear-cut approach to defining the scope, objectives, and implementation practices of ILM. Rather, a range of business objectives can serve as policy and operational drivers for ILM, and firms have adopted multiple approaches. The organization of ILM projects also varied widely according to goals and sponsorship. As a general observation, multi-function (enterprise level) ILM and data warehousing or BI efforts required relatively large sponsorship networks and executive level commitments of time and stewardship across the full planning and project lifecycles. In the words of one executive, "these projects can be significant and require the active support of top management… they (generally) involve customer data and the information necessary to meet functional or enterprise goals." Conversely, projects arising out of legal or compliance directives usually required immediate attention – "you don't want to be starting from scratch on one of these."

In interviews with users and vendors alike, most see ILM evolving beyond storage into the realm of enterprise information management. With that evolution will need to come powerful software to realize the vision. If the future for storage is in the value of storage management software, then increasingly we should see vendors enacting software strategies in partnerships, OEM deals, and software acquisitions. As users continue to apply ILM to problems they are experiencing, storage vendors will continue to broaden into software markets that will fill out their ILM product lines and implementation stories.

In a recent study of the growth of digital information, EMC/ IDC found that by 2010, while nearly 70% of the digital universe will be created by individuals, organizations will be responsible for the security, privacy, reliability, and compliance of at least 85% of the digital universe. And somewhat ominously, they advised that the cost of not responding to the avalanche of digital information growth can add up, yet not be immediately visible to CEOs and CFOs.[7]

[7] IDC White Paper, The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010 (March 2007)

VII. Postscript: Where Is ISIC's Future ILM Research Headed?

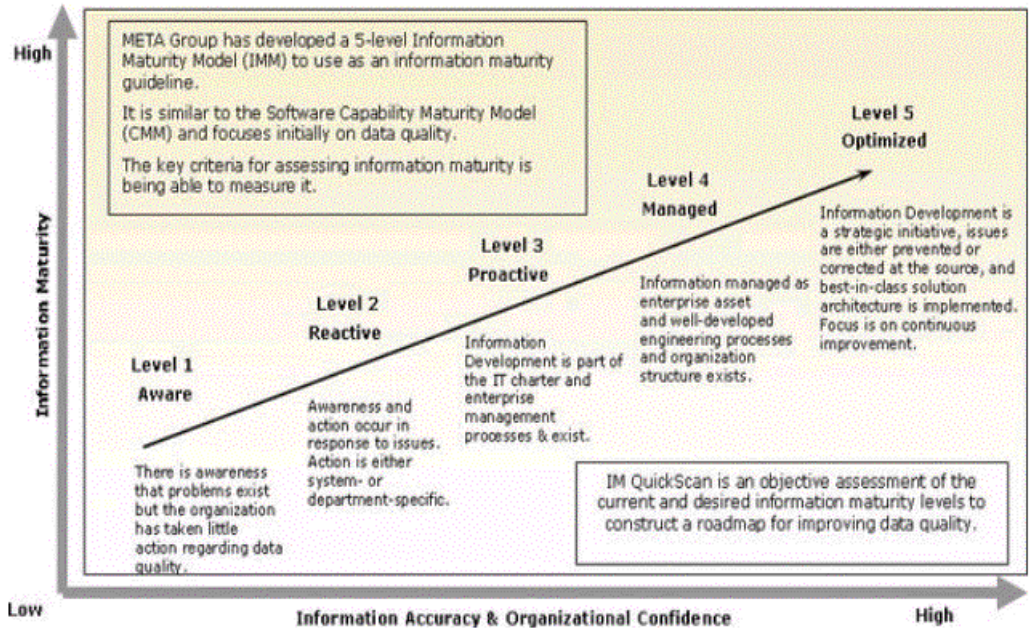
This report concludes the first phase of ISIC's ILM research project, conducted with the help, support and guidance of the SIM APC.

The next phase of ISIC's information research will involve two main areas of work. First, ISIC is undertaking a large-scale project on measuring and estimated information growth, a follow-on, replicate study of the Lyman and Varian How Much Information? studies conducted at UC Berkeley published in 2000 and 2003.[8] ISIC will study the measurement and growth of all information, with specific focus on the growth of enterprise and personal information, following the protocols used in the Berkeley studies (with extensions). The specific organization and implementation of this research will be developed over the next several months and announced in October 2007.

Secondly, ISIC will continue to study the evolution of Information Lifecycle Management, its principal components, information value and data classification, and a superset of ILM, the Information Maturity Model. Different stage models of information maturity have been proposed - these models are meant to provide an information-oriented equivalent to the Capability Maturity Model (CMM), used to assess an enterprise's software development maturity level. The takeoff to ISIC's IMM project will be a five stage maturity model defined by the Meta Group, and used in the Open Methodology Framework wiki, the MIKE2.0 Methodology (MIKE2.0 is an Open Source methodology for Enterprise Information Management). The model is illustrated in Figure Eleven:

[8] Lyman, P. and Varian, H. How Much Information? UC Berkeley School of Information Management and Systems, 2003. http://www.sims.berkeley.edu/research/projects/how-muchinfo-2003/.

Figure Eleven – The Information Maturity Model



The 5 levels of information maturity can be summarized as follows:

- A Level 1 organization has no common information practices. Any pockets of information management maturity that the organization has are based on the experience and initiatives of individuals.

- A Level 2 organization has little in the way of enterprise information management practices. However, certain departments are aware of the importance of professionally managing information assets and have developed common practices used within their projects. At the enterprise level, a level 2 organization reacts to data quality issues as they arise.

- A Level 3 organization has a significant degree of information management maturity. Enterprise awareness, policies, procedures, and standards exist and are generally utilized across all enterprise projects. At level 3, the information management practices are sponsored by and managed by IT.

- A Level 4 organization manages information as an enterprise asset. The business is heavily engaged in information management procedures and takes responsibility for

the quality of information that they manage. A level 4 organization has many mature and best-in-class practices and utilizes audits to ensure compliance across all projects.

- A Level 5 organization considers information to be as much an enterprise asset as financial and material assets. A level 5 organization has best-in-class information management practices that are utilized across all enterprise projects. The distinguishing characteristic of a level 5 organization is the focus on continuous improvement. At level 5, all data management practices and assets are regularly measured and the results are analyzed as the basis for process improvement.

ISIC's IMM research will address the feasibility of refining and operationalizing an IMM model for field data collection and analysis, initially as a benchmark diagnostic, and later as a schema for developing better and more precise measures of enterprise information across information lifecycles, and the management issues therein.

The second phase project will commence in October 2007.

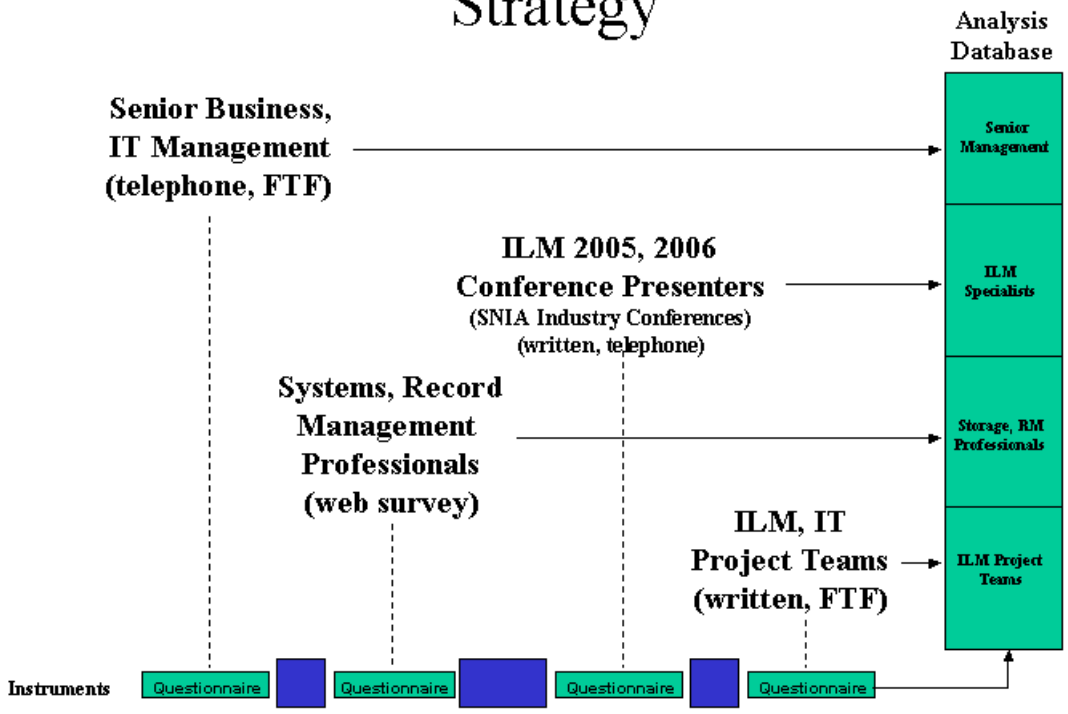
ACKNOWLEDGEMENTS

The author would like to acknowledge the support provided by Society of Information Management's Advanced Practices Council, its members, research director Professor Blake Ives, University of Houston, and Program Director Madeline Weiss. I would also like to acknowledge the time and support of executives in the fifteen case companies involved in this project, and the 345 IT and storage professionals who completed the online ILM survey.

APPENDIX

I. ILM Field Study Protocols, Survey Instruments

Figure Twelve: ILM Study Field Strategy



Four questionnaires were developed and administered to different respondent groups:

(1) a telephone questionnaire with senior executives in case study field sites
(2) an online web survey of storage and IT professionals (reproduced here)
(3) a project team questionnaire administered to ILM project teams
(4) a short form interview questionnaire used at industry trade conferences

APPENDIX (con)

Fifteen field sites participated in this phase of research. Figure Thirteen summarizes several key characteristics of the participating firms.

Figure Thirteen - ILM Case Study Sites

Industry	Firm	Size	Contact	Status of ILM, DM Initiatives	Triggering Event	Ownership	Comments
Media & Entertainment	Media1	Large	CIO	Needs Assessment	Product Innov, Cost	CIO, Line Business	
Media & Entertainment	Media2	Large	CIO/CTO	Needs Assessment	Product Innov, Cost	CIO, SBU Heads	Educating the business
Financial Services	Pension1	Large	CIO/CTO	Pilot DM, ILM Projects	Customer Experience	CIO/CTO Line Business	Partnering with business
Financial Services	Bank1	Large	Int'l Lending VP	Needs Assessment	Audit & Compliance	Audit, Line Business	Operational
Financial Services	Bank2 (Europe)	Small	CEO, CTO	Needs Assessment	Shared Services	CEO CTO Services Firm	Customer Info, Compliance
Pharma	Pharma1	Medium	Research Dir	Pilot DM Projects	Regulation, Cost, Time	CIO, Dept Heads	Pre-clinical vs Clinical
Pharma	Pharma2 (Europe)	Medium	Head of Storage	Needs Assessment	Regulation, Cost, Time		
Legal Services	Law1	Small	Founder and Partner	Procedures in Place	Litigation Success	Partners, IS	Trial preparation
Venture Capital	VC1	Small	Founder and Partner	Embryonic	Information Advantage	Partners, IS	Market assessment
Telecomm	Telco1	Large	Audit Group	Court Mandated	SEC	Corporate	
Professional Services	Consulting1	Large	CIO	Needs Assessment	Information Sharing	CIO	
Health and Medical	Hospital1	Large Bed	CIO / CTO	Multiple DM Initiatives	Patient Safety, HIS	Board, CIO, Physicians	
Government	National Archives	Large					
Government	Military	Large	Head of Logistics	Multiple DM Initiatives	Operational Readiness		
Non-Profit	Church of Jesus Christ LDS	Large		Multiple DM Initivatives	World Family History		

APPENDIX (con)

The ILM Professional Survey

Q1. Following is a list of IT topics that are often associated with Information Lifecycle Management (ILM) projects found in industry. Please rate each topic in terms of how important it is in influencing your own company's approach to ILM. 10 is the highest in importance, and 1 is the lowest.

- Tiered Storage Architecture, Hierarchical Storage Management
- Storage Resource Management
- Data Archiving
- Application Archiving
- Database Systems and Applications Integration Across Business Units
- Database Search, Data Mining, Business Intelligence
- Data Warehousing
- Data Protection, Business Continuity
- Records Management
- Compliance and Audit Systems

Q2. What other factors are influencing your company's approach to ILM?

Q3. Which of the following statements most closely matches your own definition of ILM?

- ILM is most importantly an approach to solving the problems of distributed data, database applications, and storage management in the firm, representing a natural progression from storage area networks, tiered storage, and hierarchical storage management (HSM).
- ILM is most importantly a policy-based approach to improving records and information management within the firm, including defining ownership of data, patterns of data access and use, data security, and data archiving
- ILM is most importantly a business application within data warehousing and business intelligence, supporting the linking of warehouse data with customer relationship tools and analytics.

APPENDIX (con)

Q4. What do you see as the main benefits of ILM?

Q5. What do you see as the main drawbacks of ILM?

Q6. Are you currently involved, or have you been involved in, an ILM-related project within your company? If Yes go to Q7. If No go to Q9.

Q7. Please briefly describe the project and your role in it.

Q8. What are the main goals of the project?

Q9. Are you a/an:
- IT/IS Professional – Storage User
- IT/IS Professional – Storage Vendor
- Systems Integrator / VAR / VAD /
- Consultant
- Records Manager
- Analyst
- Member of the Press
- Trainer/Educator/Researcher
- Student
- Other (Please Specify)

Q10. Which of the following best describes your current job?
- Consultant
- Design/Development
- IT Management (Manager IS/IT)
- Network Management/Administration
- Non-IT Staff
- Project Management
- Research
- Sales/Marketing
- Senior IT Mgmt (CIO, CTO, Director)
- Senior Non-IT Mgmt (CEO, CFO, Director)
- Student
- Systems Management/Administration
- Records Management/Administration
- Other (Please Specify)

SELECTED BIBLIOGRAPHY

"An Approach to Data Classification in an ILM Framework," SUN Microsystems White Paper, October 2005.

Aral, S., Brynjolfsson, E., and Wu, D.J., "Which Came First, IT or Productivity? The Virtuous Cycle of Investment and Use in Enterprise Systems," 27th Annual ICIS Conference, Milwaukee, Wisconsin 2006.

ARMA International (2005). "Information Management: A Business Imperative." Available at www.arma.org.

Ballou, D., Madnick, S.E. & Wang, R.Y. (2004). "Special Section: Assuring Information Quality." Journal of Management Information Systems, Vol. 20 No. 3, Winter 2004 pp. 9-11.

"Best Practices in Information Lifecycle Management for Private Sector Compliance," SUN Microsystems White Paper, October 2005.

"Best Practices in Information Lifecycle Management Security," SUN Microsystems White Paper, February 2006.

Brohman, M.K., Parent, M., Pearce, M.R. & W. Wade (2000). "The Business Intelligence Value Chain: Data-Driven Decision Support in a Data Warehouse Environment: An Exploratory Study." Proceedings of the 33rd Hawaii International Conference on System Sciences.

Casassa-Mont, Marco, "On Privacy Aware Information Lifecycle Management (ILM) in Enterprises: Setting The Context," Hewlett Packard Labs, Information Security Solutions Europe (ISSE), 2006.

Chen, Y. (2005). "Information Valuation for Information Lifecycle Management." Second International Conference on Autonomic Computing (ICAC), IEEE Computer Society, pp. 135-146.

Cook, R. (2003). "Here Comes ILM." SearchStorage.Com, July 21, 2003. Available at www.searchstorage.com.

Duplessie, S., Marrone, N. & S. Kenniston (2003). "The New Buzzwords: Information Lifecycle Management." SNW Online, ComputerWorld, March 31, 2003.

"Emerging Opportunities in Database Information Life-Cycle Management," META Group, Stamford Ct (metagroup.com), January 2005.

Gable, J. (2005). "IT Update: What CIOs Should Know About Records." ARMA International (www.arma.org).

IDC, The Expanding Digital Universe: A Forecast of Worldwide Growth Through 2010. International Data Corporation, March 2007.

"Information Lifecycle Management: An Automated Approach," EMC Technical White Paper, December 3, 2005.

"Information Lifecycle Management for Business Data," Oracle White Paper, Sept 2005.

"Information Lifecycle Management Maturity Model," SUN Microsystems White Paper, April 2005.

"Information Lifecycle Management Vision," SUN Microsystems White Paper, December 2005

Kaarst-Brown, M.L. & S. Kelly (2005). "IT Governance and Sarbanes-Oxley: The latest sales pitch or real challenges for the IT Function?" Proceedings of the 38th Hawaii International Conference on System Sciences.

Knorr, Eric, "Building the Compliance Infrastructure: Service-oriented architectures have found their way to the network," CIO Magazine, July 15, 2005.

Komiega, K. (2004). "ILM Standard in the Works." SearchStorage.Com, April 8, 2004. Available at www.searchstorage.com.

Lawyer, J. & S. Chowdhury (2004). "Best Practices in Data Warehousing to Support Business Initiatives and Needs." Proceedings of the 37th Hawaii International Conference on System Sciences.

Lyman, P. and Varian, H. How Much Information? UC, Berkley, School of Information Management and Systems, 2003. http://www.sims.berkeley.edu/research/projects/how-muchinfo-2003/.

Madnick, S.E., Wang, R.Y. & X. Xian (2004). "The Design and Implementation of a Corporate Householding Knowledge Processor to Improve Data Quality." Journal of Management Information Systems, Vol. 20 No. 3,Winter 2004.

Peterson, M. (2005). "Strategic Profile: Information Lifecycle Management: A Vision for the Future." Storage Networking Industry Association (SNIA), Data Management Forum.

Senf, D. (2005). "Building Towards An ILM Strategy." International Data Corporation (IDC). Available at www.idc.com.

SNIA Data Management Forum (2005). "Information Lifecycle Management: A Vision for the Future," Storage Networking Industry Association.

Tallon, Paul, "The Discipline of Managing Storage Costs: A Research Perspective," Glasshouse Technologies Whitepaper, 2006

Tallon, Paul, "The Viability of Indefinite Storage: Insights from Email Archiving," Glasshouse Technologies Whitepaper, 2006.

Ulfelder, Steve, "Information Lifecycle Management Lives, Finally," Application Development Trends, August 1, 2005. Weblink: http://www.adtmag.com/article.aspx?id=11501&page=

Weill, P. and M. Broadbent, "Leveraging the New Infrastructure: How Market Leaders Capitalize on IT," Harvard Business School Press, 1998.

Wendt, J.M. (2003). "Starting the ILM Process." Storage Magazine, December 2003. Available at http://storageMagazine.techtarget.com/magItem/0,291266,sid35_gci941220,00.html