

Managing Your Organization's (Information) Assets

By James E. Geis, Forsythe

With all the industry buzz around ILM (information lifecycle management), tiered storage, SRM (storage resource management), DR (disaster recovery), BC (business continuity), backup, continuous data protection (CDP) and restore, and the myriad of other acronyms for the management of information, effective information storage management really boils down to a company's definitively understanding and paying attention to the four phases of its information's metamorphosis: Creation, Access, Retention and Deletion ("CARD"). Only by implementing policies, procedures, and technologies to effectively handle information through all four phases can an organization ensure that its critical information assets are treated in a way that storage is secure and recoverable. Likewise, through understanding its "CARD" phases, an organization can reduce its investment in the storage infrastructure supporting the recoverability of all its essential and non-essential information.

Many vendors, industry and end-user groups, and standards organizations are currently defining the mindset for managing information and refining the concepts that make up information lifecycle management. Leading the charge is the Data Management Forum of the Storage Networking Industry Association (SNIA). This group brings together participants from the end-user community, as well as storage and information management software vendors, and other records management organizations like the Association of Records Managers and Administrators International (ARMA). The end result of their efforts will be a standard framework within which all information management concepts converge. But, as with any standard, it will be some time before it's fully fleshed out, reviewed and revised, agreed upon, widely accepted and adopted as a standard, and integrated into software tools. And once it becomes a widely-accepted standard, experience suggests that many vendors will add their own individual nuances, much as has occurred within open systems technologies. But the need for a standard is undeniable: end-users are demanding it, businesses can't function without it, and the industry can't sustain the lingering confusion any longer.

Even with the adoption of such a standard, however, individual organizations will still need to understand the four

basic phases of information metamorphosis as they apply to their own corporate and customer data. Only once the information creation, access, retention, and deletion phases have been outlined, documented, and communicated, and the transition between each phase identified, can an organization choose the appropriate storage technology, software, and operational processes for its short- and long-term information storage needs. Each type of information in each phase has varying service and operating level requirements because of differences in usage patterns. Each also supports different business requirements. These requirements dictate the business- and compliance-driven long-term storage, disaster recovery, and restore requirements, which in turn dictate the appropriate ITIL best practices and the optimal logical and physical storage structures.

Developing a "CARD" (Creation-Access-Retention-Deletion) Mindset

Creation and Access are usually pretty clear cut; most organizations know (or should know) why information is created and why and by whom it is supposed to be accessed. But, understanding or defining Retention and Deletion requirements can be a Rorschach test for many companies, as some regulations leave them up to organizational policy. Some organizations have a "retain forever" policy, while some have stringent destruction requirements. A policy at either end of the spectrum is taxing for a storage or IT administrator, as well as for the legal department and other groups that may need to locate information on demand within a certain time frame.

Creation Phase

Clarifying all four phases is best done by starting at the beginning, and understanding all the variables—the who, what, where, when, why, and how? of the creation phase for all information. Understanding these facts is essential to successfully calculating the value of each type of data and, therefore, its appropriate treatment. Unfortunately, value can't simply be calculated. However, in order to accurately prioritize each type of information, it is imperative to assign a relative value to it, through careful consideration of the above-mentioned variables.

Accurately prioritizing data also requires the assignment of ownership and accountability. These are usually application- or business-unit based. Generally, when information is dynamic or “fresh”, its associated “value” is perceived to be higher than when it becomes static or reference information. Aside from its internal value to the organization, external regulations can impute a value of their own. For any piece of information, your organization must decide what fiscal, legal, historical, and administrative requirements it fulfills.

Once the business value of the data has been determined, it is possible to make prudent, strategic decisions about the amount of investment that can and must be made in its handling throughout all four phases. For example, compliance regulations can dictate requirements for performance, recoverability, and number of available copies. Regulatory requirements, combined with business requirements, can then be translated into policies, procedures, and technologies, the cost of which is weighed against the requirements. The actual cost of information management encompasses many variables, including physical storage, hardware and software management, liability, not to mention the personnel resources involved through all four phases. For every primary byte of data stored, there is a multiplier attached as a result of multiple “touches” to the information and multiple copies that may exist. At some point this multiplier may become exponential and often unmanageable as a result of replicated copies on disk or tape. Reducing this multiplier is the argument behind ILM.

Access Phase

During the access phase, value, importance, security, encryption and recoverability requirements most likely will not change. However, it is frequency of access that dictates physical or logical location. During the access period, the performance requirements usually change, which may require a change of platform or location from where the information was first created. But the application will still need to locate the information. This is where ILM fits into tiered storage. The decision point comes to, “At what exact point (time, value or usage) does the information transition from active to inactive or from temporary to permanent or from dynamic to static, and how is that determination made?” The access period can extend well into the retention period, which can be for decades, hence, the question of long-term storage on disk versus tape. The options for storage protocols and applications for information access are becoming more abundant, but also more diverse. As much as the variety of possibilities offers, it complicates storage and information management; hence, the focus on standards. In turn, such technologies are not always easily integrated into the infrastructure. For example,

as storage converges more towards the IP network, security and accessibility are becoming even more complicated.

Retention Phase

This is one of the more difficult phases that many organizations struggle with. How long must information be kept? In some industry-verticals subject to specific regulatory guidelines, such as healthcare and finance, it’s pretty clear. But the onslaught of other “general” corporate regulations, as well as pending national privacy legislation will affect technology and process decisions for the long term. Every user group or application owner is going to feel that their information is “the most important” to the organization until they see the price tag. So who is the final arbiter? Is it legal, finance, or the information-owner? Having the ability to store and find the needle in the haystack is what standards-based storage organizations, and the storage industry, are striving for. Everyone wants the elusive, all-encompassing index/search/find/retrieve engine.

How long an information element must be retained dictates technology decisions (disk, tape, optical, or paper storage?). The operational processes to support retention must be instilled and married to technology. Having a schedule with all information types, locations, owners, and retention periods is prudent. The long-term technology is the other difficult factor, as the technology acquisition and integration curve has sharpened, and the terabyte disk is just around the corner. Will the technology you purchase today work with the software and applications of tomorrow?

Deletion Phase

Don’t hit that delete button just yet. Sometimes, it’s a matter of not deleting the information. A large factor in the retention phase that leads into deletion is preventing the destruction of information when legal, or other circumstances dictate. A new word in the IT vocabulary is “spoliation.” It is the legal term for the intentional alteration or destruction of a document. So, in trying to determine when you can delete a piece of information, be sure to ask your legal group—and don’t give up until you get an answer, because they may someday have to stand up in court and explain why or why not something was deleted. A related question is what determines permanent archive? This eliminates the need for some challenging decisions, but the longer you hold off on the delete button, the higher the cost of retention. So be sure you know what legal, fiduciary and administrative reasons allow deletion?

And when you delete, are you sure that ALL copies of the record have been destroyed on all replicas (tapes, media, personal computers, PDA)? And can you prevent destruction if required to do so? Stop and ask your organization these questions: Does your organization have a documented

records retention schedule that includes the type of record, retention period, location(s), access requirements, and continued security paradigms in place? Have you determined the least (or highest) common denominator when it comes to access, retention, and deletion, and do you have the mechanisms in place to halt destruction for legal reasons (in litigation) or disgruntled employees or simply due to user error?

Understanding “CARD” Strengthens Information Management Policy

This core of this exercise is that each and every phase of CARD must have an associated policy. In some instances the responsibility or decision point must either be removed from the end-user or automated. Every policy must have a documented procedure, and each transition from phase to phase must be documented and assigned an owner. Even more important is putting in place a communication and training strategy, audit and monitoring abilities, and someone responsible for checking the process to ensure that policy is followed.

Technology has simultaneously complicated and simplified our jobs as information managers. Punch cards suddenly seem like an attractive alternative. Businesses run 24x7 and have reputations at stake. Information is an asset, and the responsibility crosses many boundaries within the organization. We’re still quite a bit away from the having a centralized view into all information, but until then, understanding the lifecycle phases can help keep policy, technology and operational decisions aligned with the businesses overall needs.

Published in Computer Technology Review, March 8, 2007.

James E. Geis is director of integrated solutions development for Forsythe. Geis developed Forsythe’s unique information management framework—the roadmap Forsythe uses for information management, storage and server consulting engagements. He currently is focused on developing solutions to address the impact of virtualization on computing technology and the data center.

