



Contemporary Corporate Regulatory Legislation and its Impact on Information Technology

by Alan Stuart,
Chief Strategist, Content and Data Retention Solutions,
IBM Software Group

ABSTRACT:

Organizations today are struggling to understand and to comply with regulations requiring the retention and preservation of many forms of business records, in particular, electronic records. Although a strategy to attack the problem, then define and implement the solution might be stated quite simply it is generally not so simple when one tries to implement. In its most basic form, a compliance strategy comes down to being able to answer three questions:

1. What records does my organization need to preserve and for how long? Typically, input is required from legal, audit and various consultants.
2. What software/middleware and services are required to implement my organization's specific records management requirements? Typical solutions will employ various forms of middleware, such as content management, records management, and storage management.
3. What storage solution does my organization need to preserve our corporate records during their retention period? Typical solutions employ a wide range of storage devices, disk, tape, optical, DVD, and use new data protection options, such as retention managed storage, to ensure the integrity of the data.

BUSINESS ENVIRONMENT – THE TREND TOWARD RETENTION

In speaking to technology executives about regulatory compliance, one hears the story of the three bears again. They ask, "Am I doing too much? Am I doing too little? How do I know when we have it just right?" Does this sound like déjà vu? These same questions were being asked prior to Y2K. The current rush by organizations to achieve compliance with the plethora of federal, state, and local regulations is certainly reminiscent of the rush to be Y2K compliant. However, the difference here is that regulatory compliance is not a one shot occurrence. It is something that every organization has to weave into the fabric of their business operations. It is a process that will evolve over many years. Failing to find that 'last bug' from Y2K may have caused an organization to experience a small operational outage. Yet, failing to comply properly with some of the current governance regulations that are now in effect (or about to go into effect), can cost a company millions of dollars in fines, damage to their business reputation and now, even the possibility of jail for the responsible executives. This may turn out to be the case where 'good enough' is not quite bullet proof enough to let business and IT executives sleep peacefully. Typically an organization obtains professional guidance from multiple sources, including lawyers, accountants and auditors, in order to determine where they are and what is required to achieve regulatory compliance and manage risk for their particular situation.

As the conversation continues, the executives describe a number of different approaches to solving the retention and preservation problems. Some say that, if they are required to keep the data for three years, they will have it deleted in three years and 10 minutes. Others say that in today's litigious environment the data they archive could be used for their defense in the case of lawsuits and auditing inquiries. These executives have decided to keep all of their data forever. Many other executives simply say they are trying to figure out what policies they need to establish and how these policies will effect their operations and their budgets.

It is important to note that in enforcing and monitoring corporate compliance with regulations, regulatory agencies have two constituencies, the public and the legislative body that enacted the law, i.e. The U.S. Congress. A regulator's worst nightmare is that an organization followed the regulatory requirements to the letter and yet there was still enough wiggle room to cause another highly visible corporate scandal. So, the screws are being tightened. Regulatory oversight is now much more intense. Legislators around the world have been very busy passing new laws which require companies to do things that, in fact, might improve their business in the long run. Furthermore, these regulations and their interpretation are continuously evolving.

By passing legislation such as the Sarbanes-Oxley Act (SOX) of 2002, Congress indicates to organizations that ignorance isn't bliss; it is illegal. Taking away deniability from corporate executives and imposing the threat of heavy fines and jail would be treated as a very serious matter. However, if that is not enough, the legislation goes further to protect "whistleblowers" in the event that the spirit and intent of SOX does not get to everyone in the Board Room.

So what does a company have to do to be SOX compliant? Most importantly companies should identify a set of consultants, legal advisors, and domain experts to track and advise the most appropriate and current interpretations. In its simplest form, SOX, and

other similar legislation like the Gramm-Leach-Bliley Act of 1999 (GLBA)[1] impose an 'ISO9000-like' rigor to corporate business processes, record keeping, and information management. It tells companies that they need to document their business processes, that everyone up and down the management chain needs to be aware of these processes, and that the documents associated with these processes need to be kept for a period of time. Moreover, an organization needs to assure that their business records are not tampered with during their retention period.

What an organization needs to do is to take a "Back to Basics," business management 101 approach to this challenge. This requires taking an integrated view of process, organization, technology and controls:

- Processes-- Efficient, Effective, Timely, Standardized, Systematized
- Organizations and People-- Skills, Capabilities, Understanding, Responsibilities, Accountabilities, Governance, Culture and Values
- Technology and Systems-- Consistency, Data Integrity, Embedded Workflow and Controls, Security, Optimization, Integration
- Controls-- Documented, Preventative, Detective, Embedded, Integrated, Aligned, Measured, Monitored

Having procedures, policies and controls that are effective, efficient, well run, and well documented will facilitate an organization's efforts to achieve SOX compliance and at the same time have the opportunity to make improvements in business performance.

In addition, an IT solution is needed when implementing the records management policies to preserve records and to insure the integrity during the required retention period.

FOCUSING ON CONTENT AND DATA RETENTION

Understanding what records need to be kept and how for long is very important. However, without the appropriate IT solution, storing, preserving and retrieving the records can be an almost impossible task.



In October 2003, IBM's Business Consulting Services organization completed a SOX readiness survey. Executives were interviewed about their readiness for SOX and the challenges they faced. According to the survey, they consider records management, IT infrastructure, and accountability issues as the greatest challenges in seeking SOX compliance. Many IT organizations already have some combination of content management, records management, and storage management middleware in house supporting the retention and retrieval of a wide variety of business records. Many organizations are extending existing policies and procedures to this new record class with new retention policies. Some are adding a records manager to enhance existing content management capabilities.[2] Some organizations are starting from scratch. No matter where an organization is in this process, there are many products and services available to help implement a solid IT infrastructure for a content management application. However, there are new features that the middleware should include in order to effectively manage records for regulatory compliance.

Event Driven Retention: In addition to normal retention period strategies, there are event-driven retention requirements. While some records can be identified and retained for a fixed time period (e.g. save this e-mail for three years), other records, for example, a consumer's application for a brokerage account, be retained for some unspecified period of time, in this case, until the account is closed. Further in event-type retention, there may be a minimum retention period and/or a specific retention period after the event has occurred. For example, after a consumer closes a brokerage account, keep the data for six years.

Audit/Legal/Deletion Hold: As some companies have found out recently, the government frowns upon the destruction of records after a legal action has commenced. So a second new requirement is that there needs to be a capability to put a 'hold' on a set of records that are the subject of an audit or legal proceeding. Remember the executive who said that he would have the data deleted in three years and 10 minutes? His company is currently in court. It seems that the company received a subpoena about some set of transactions just before the regulatory retention period expired. However, the procedures in his legal department took a little over two months to request the records from the IT Department. By the time the IT Department received the request from Legal, all of the data and all of the backups had already been deleted. To help avoid a situation like this an organization could:

- improve the business process between legal and IT;
- add some fixed latency to the required or planned records retention period prior to deletion; and
- implement the technology to provide a simple way to place a deletion hold on the records.

Our executive who proposed deleting records 10 minutes after their regulated retention period expired might want to rethink that strategy.

These two new requirements need to be implemented and integrated in the records manager, content manager and storage manager. The content manager middleware needs to be able to identify event based records, signal when the event occurs and allow for minimum retention and retention after the event occurs. This requires very tight integration between the content manager and the storage manager. If a record manager is being used, the tight integration applies to all three components. A piece parts solution here may not yield the desired results.

Besides the new middleware functions described above, managing a repository of records for regulatory compliance requires an organization to rethink their backup and operational recovery strategies. Many organizations use the capabilities of a storage system to provide disk-to-disk remote copy as their backup/operational recovery mechanism. Other organizations use conventional backups, periodically taking a complete copy of the archive and then using incremental backups in between. Let's look at these two approaches in the context of an archive for regulatory compliance.

Disk-to-Disk Remote Copy: Disk-to-disk remote copy is commonly used practice for backup and recovery of regulatory records; however, it is sub-optimal. In the case of a failure, the content manager application receives an error message to alert an operator, who will then execute the storage device's recovery procedure to recover the 'damaged' record. Then, the application can attempt to retrieve the record. Further, when the content manager rightfully requests that the record be deleted, how do you insure that the remote copy has also been deleted? Is this an opportunity for a rogue backup to be found by a forensic program in a litigation? A better way to approach this is to have the content manager application create as many copies as needed, in as many locations, when the record is first stored in its primary archive. This provides two benefits. First, in the event of a read error, the content manager already knows where the backup copies are and can seamlessly access the backup copy and complete the transaction.^[3] Second, the content manager provides referential integrity that helps to insure that all copies of the record are kept in sync. Further, as discussed in the next section, many implementations of a regulatory archive require multiple copies of records in multiple locations. A single point-to-point backup done by a storage device may not provide adequate protection for these kinds of records.

Traditional Periodic Backup/Incremental Backup: As with disk-to-disk remote copy, traditional backups are an acceptable method, up to a point. A full backup is taken, and then incremental backups are taken for some period of time. Then a full backup is taken and the cycle starts again. This can become a very time consuming process as the repository grows from 5 to 10 to 30+ terabytes or larger. Further, while some regulations require at least one remote copy, other regulations (or corporate policies) require two or more backup copies be maintained. A more effective way of managing a large repository that requires multiple replications is to use the content manager to make as many replications of the record at the time the record first enters the repository. For example, in the financial services industry, there are check-image archiving applications which manage incredibly large repositories. These "mega archives," which approach 4PB in size represent some of the largest archives on Earth. Each day, these particular archives currently receive up to 60-80 million checks, generating 1.5TB of data. The content manager creates four backup copies of each check image at the same time the primary copy is stored. If this was not done at creation, it would be impossible within batch timeframes to create the four backups using any traditional method. If you are using traditional backups today, changing to this method could eliminate a significant amount of work.

When you consider both of these backup and recovery scenarios together, implementing them may yield a significant operational improvement.



TECHNOLOGY ENVIRONMENT – THE TREND TOWARD PRESERVATION

Preserving records for a long period of time (years) also adds a new dimension to selecting the storage to house the data. In order to store and manage records more than a year, a solution should:

1. Preserve the bits. Maintain the integrity of the records. This is a function of the reliability of the media and the robustness of the system.
2. Migrate the bits. Depending upon the media used to store the information, and the length of time retained, there may be a need to migrate the records from older media to new.
3. Interpret the bits. After all that time preserving and migrating bits, at some point in the distant future, someone may want to view, mine, or process the information. You need to plan for this part of the life cycle at data capture.

While today's business environment requires data to be retained for certain periods of time, the technology environment is required to preserve the data for the duration of the retention period. The impact of this increased focus on business process,

business controls, and organizational issues has also increased the need to retain large quantities of data for longer periods of time.^[4] Taking a storage-centric view, this class of data is being called fixed content data, archive data, reference data, unstructured data, and many other terms which imply its write-once, read many (WORM) characteristic. However, the most important management attribute of this kind of data is its retention period, hence it can be called Retention Managed Data. Today, there is a lot of vendor hype around this kind of data. Some organizations don't know where to start, but they know managing this data is a key component of their regulatory compliance. Further, Section 802 of SOX states, "Whoever knowingly alters, destroys, mutilates, conceals, covers up, falsifies, or makes a false entry in any record, document, or tangible object with the intent to impede, obstruct, or influence the investigation... shall be fined under this title, imprisoned not more than 20 years, or both." As organizations determine how they are going to respond to this part of the regulation, they are looking toward strengthening their records management and retention policies with policy-based storage management that can help insure the protection and integrity of these records.

As a result, there is a special category of retention managed data that is compliance data or data whose protection is required by some regulatory guidance.^[5] Some regulations specify that the storage system provide an enforced non-rewriteability and non-erasability for all the data retained under the regulations. Many refer to these attributes as Write Once Read Many (WORM). However, using the term WORM in this context can be confusing.^[6]

What is the right storage media for a regulatory archive? There is one correct answer to that question -- "it depends". It depends on many factors:

The size of the archive; number of records; size of each record – Is this an archive of billions of relatively small records or thousands of very large records (this will effect the next factor)

1. Data protection – Do you need to (or want to) protect the data from modification and/or deletion via specific storage media or storage management controls? If so, how many copies of the data need this protection?
2. Life cycle – Depending upon the length of time since the record entered the archive:
3. Length of retention period-- How long is this record going to be managed?
4. Frequency of access to record-- Data archived for regulatory purposes is rarely accessed and the likelihood of access drops to near zero as time goes by.
5. Read performance-- What is the latency requirement for accessing a record? Once accessed how much bandwidth is required?^[7] The need to access a record quickly diminishes over time. The frequency of customer statement inquiries drops after 90 days, after 12 months and to near zero after two years (Individual application and organizations will have different break points). The point is that once the frequency of access drops very low, there is an opportunity to move the data for the rest of its life to a less expensive media (i.e. tape).

Selecting just the right mix of storage that meets the technical requirements and affordability requirements is a juggling act. The correct answer is -- "it depends".

SUMMARY

Sarbanes-Oxley, and other similar regulations, have caused a bit of a feeding frenzy between vendors and organizations. Many vendors are out there today selling the point solutions they have in their sales kit. Some vendors are acquiring companies to make up for the lack of end-to-end, integrated capabilities that are required to efficiently provide a solution. A few vendors have all the components to satisfy a wide variety of requirements. The solution is not a storage box and it is not a piece of software/middleware. A complete end-to-end solution needs to be considered. In this environment it is a good thing to find a single vendor who can provide a complete solution; however, it is also not necessary to acquire all the services, software and hardware from a single vendor. However, the solution needs to be integrated so that all the components work well together. To put a proper solution in place, it still goes back to the basics: understand what records need to be kept and for how long, implement a software/middleware solution which satisfies the retention requirements and purchase storage which meets the preservation requirements in a cost effective manner.

^[1]GLBA focuses on information security and privacy in the financial services industry.

[2]Record management middleware has sophisticated algorithms to determine the correct retention period and disposition for each record.

[3]It is also expected that the content manager will raise an alert of some kind so that the primary record is restored.

[4]The large repository sizes and long retention periods require a significant change to the way we compute total cost of ownership (TCO) is calculated. When you consider the three key factors, (1) initial storage acquisition is larger, (2) length of time retained is larger, thus raising the possibility of having to replace old media, and (3) power and cooling costs for (a) larger storage and (b) over a longer period of time becomes significant. In looking at these new dimensions, tape (because of its lower cost), and a multi-tiered storage architecture, becomes a compelling part of a regulatory archive repository storage strategy.

[5]Refer to The Sarbanes-Oxley Act of 2002, Section 802 for specific wording and penalties.

[6]The confusion can arise from the point of view. Is the data WORM? Is the media WORM? Is there some control code that enforces WORM-like retention management policies? The term WORM is appropriate when referring to the media if the media is non-rewriteable and non-erasable, such as WORM Optical and now WORM Tape. However, the current technology for providing these attributes on a naturally rewriteable disk media is an in-band software/control code/microcode solution. Hence, using the term "WORM disk" is not correct in this context because the disk media itself does not provide the non-rewriteability and non-erasability. A better term for this kind of storage is retention managed disk storage. This is because the enforcement of these storage qualities is in the in-band control code, such as IBM's Tivoli Storage Manager for Data Retention or EMC's CentraStar software, not on the disk media.

[7]Write performance requirements are usually a factor of the number of records to be stored (and their size) and the arrival rates of these records.

***Datatrend's TrendSetter eNewsletter
January 15, 2004***