IMPLEMENTING DIGITAL DOCUMENT ARCHIVING FOR MEETING ORGANIZATION COMPLIANCE

Mark W. Thomas

Indiana State University
Terre Haute, USA

The recent amendments to the Federal Rules of Civil Procedure (FRCP) have organizations scrambling to assure that they have sufficient and timely solutions for archive and retrieval of digital information. Organizations are required to be continuously aware of the electronic data they have, and now must be able to retrieve the information quickly.

A June 2007 study produced some interesting and disturbing statistics about the companies that were surveyed. Per the study, more than 69 percent of companies in the United States are ill prepared to respond to litigation (Firms Unprepared for E-Discovery, 2007). Only 6 percent reported that they could quickly and confidently reply to an E-Discovery request (Firms Unprepared for E-Discovery, 2007). The study also found that most organizations have a large gap between the Information Technology function and legal departments.

Another recent survey performed by Osterman Research found that one in four businesses have had to delay a business or Information Technology initiative in order to respond to legal discovery (Firms Unprepared for E-Discovery, 2007). The survey also produced other interesting statistics about the E-Discovery dilemma. For each discovery request, 58 percent of the Information Technology department time is spent gathering, restoring, and searching backup tapes (Firms Unprepared for E-Discovery, 2007). Information Technology departments accept the majority of the costs of recovering data for legal discovery. On average, the Information Technology departments assume 62 percent of the total costs involved (Firms Unprepared for E-Discovery, 2007).

The 2006 amendments to the Federal Rules of Civil Procedure are significant because they loosely defined electronic records as another type of discoverable information. Among other things, the amendments represent a rejection to arguments that electronic information is a subset of general documents. The amendments broaden the scope of discoverable information to include digitized data such as voice mail, surveillance video, instant messages, and electronic mail. While broadening the scope of discoverable information, the amendments fail to provide a significant definition for electronically stored information. Instead, technology professionals are left to apply an inclusive approach that confirms that system data is a relevant target for E-Discovery.

The following will provide a framework for planning for an implementing a collection of policies and procedures to meet the needs of organizational compliance, e-discovery requests, and other regulatory agency requirements. The framework will include discussions on classifying data, data governance, information lifecycle, and retention and destruction.

DATA GOVERNANCE

In order to discuss the use of data, it is necessary to introduce the concept of Data Governance. Data Governance is usually defined by the presence of an executive level board or committee that creates, implements, and enforces policies and procedures with respect to the use and management of data throughout the organization (Karp, 2006). The Data Governance concept is usually tailored to meet the needs and abilities of the organization. It is an inherently cross-functional approach that engages a mixture of professionals. Data Governance doesn't govern data directly. The approach is governing how data is accessed and used during the information lifecycle (Karp, 2006). Data Governance is more about combining people, policies, and procedures to produce a larger scale process. The foundation of Data Governance is rooted in compliance, integration, and transformation. Many organizations investigate data governance programs handle diverse compliance issues involving internal policies, regulatory legislation, or standards (Karp, 2006). Some organizations look to data governance programs to govern a variety of integration technologies or manage change.

DATA CLASSIFICATION

Data classification is essential in any environment where confidentiality, integrity, and assurance are important. In order to effective implement a digital archiving policy for confidential data; data classification procedures must include marking data with role based access restrictions (Erlanger, 2005). Data should be reviewed on a periodic basis and classified according to its use, sensitivity, importance, and retention period. Typically, an organization begins with three levels of classification and adds sub-categories based on properties such as retention period, importance, and exceptionality.

The lowest level of classification would be public. Public information is considered unclassified in that it is does not contain a marking. Public information does not have special access restrictions. Although public information is not marked for advanced access protection, it does not absolve the organization from including retention and destruction schedules or archiving to ensure regulatory compliance.

The second level of classification is confidential or secret. Confidential data must be marked in order to employ role based access restrictions. Typically, confidential data would not expose the organization to loss, but should be protected to prevent unauthorized disclosure.

The third level of classification is the highest level in terms of security and retention. This level can be referred to as top secret, high risk, or restricted. The data is marked, often numbered, and access is audited. This level of classification includes data for which there are legal requirements for disclosure, privacy, retention, and E-discovery.

Data classification is an ongoing process of reviewing data and categorizing it into classification or sub-categories based on the needs of the organization. Each information source within the computing environment should be categorized, protected, and archived according to the requirements of each classification. It is important that the organization be consistent with classification in order to ensure that replicated or copied data retains its classification throughout the enterprise.

INFORMATION LIFECYCLE

The birth of a document or data record begins when the data is collected and the document created. Upon birth of the record, an owner is assigned who is responsible for classification, marking, and initial security. Data owners must determine the classification to ensure that the data steward or custodian is protecting the data in an appropriate manner relative to the classification. Therefore, at birth a document is created, classified, marked, secured, and organized in a logical system of retention and destruction (Erlanger, 2005). It is important to highlight that data may have many stewards, but only one owner. The solicitor of the information is generally considered the owner and responsible for proper handling. Data owners cannot outsource their responsibility to that document or parcel of data (Erlanger, 2005). At this stage of information lifecycle it is always important to question the need for collection of data and the creation of electronic documents.

The next stage of the lifecycle is use. Information can be processed in many different ways including printing, transferred, transformed, or copied. Data will inevitable be backed up or archived. It is important for the data classification and marking to flow with the data as it is utilized.

The final stage of the information lifecycle is death. A retention and destruction schedule should describe the how long common data types should be retained. When the time has expired, it is important to follow the destruction policy and destroy all copies of the information.

RETENTION AND DESTRUCTION

Federal regulations make it a crime to alter, falsify, or destroy documents with the intent of impeding or obstructing an official proceeding (Sarbanes-Oxley Act of 2002, 2002). In accordance, the retention and destruction policy should provide for systematic review, retention, and destruction of organization documents. The retention and destruction policy should cover all records and documents, regardless of physical form. The policy will create guidelines for retention periods and defined acceptable destruction methods.

BEST PRACTICES

The concept of best practices is ever-present in the technology industry. In the most basic sense, best practices are recommendations regarding processes or techniques for the use or implementation of products or services. Many organizations view best practices as an integrated system or approach to continuous improvement of process and procedures (Thompson, 1995). It is essential to make the distinction between best practices and common practice. Best practices should be viewed as practices that have been consistently proven to provide the desired outcome with efficiency.

The following list of best practices for digital archiving are widely accepted but may not

be universally utilized (Xiaomi, 2003). Organizations must decide which practices apply to their environment. Certain situations or will dictate the necessity to modify or ignore the practice.

Engage Stakeholders

As with any successful implementation, an archiving implementation project should solicit input from a variety of stakeholders. An understanding of the intent of the system and its likely sources will help create a bridge between functional departments.

Define needs and objectives

Understand the needs and objectives of the organization. Determine the data classification scheme, identify the regulations that must be met, and define the business needs of historic data retention. An appropriate tactic would be to interview individuals from a cross-section of the organization.

Define policy

The policy should cover all individuals or entities that come in contact with the organizational data. It should manage the retention, storage, and disposal of electronic records. Policies should be reviewed annually and modified accordingly. The policy should include definitions, security issues, responsibilities, and auditing processes.

Consider search and marking requirements

One of the key benefits of an archiving solution should be its ease of use with regard to performing searches. Searching is enabled by the use of indexes and enhanced by the marking capabilities of the system. If e-discovery or regulatory compliance is the main requirement of the archive, the solution will need to have flexible, robust, and dynamic search capabilities.

Automate the process

Employees will not have the time, expertise, or desire to manually classify and manage records. Generally, manual processes lead to incomplete and inconsistent archives. Automation will ensure uniformity and increase productivity with respect to the archiving solution. Automate the processes that benefit from automation while leaving the flexibility to override retention schedules in the event of a litigation hold.

Continuous process improvement through monitoring

Implement periodic monitoring and audits to ensure the archiving solution is providing the maximum benefit. Use continuous process improvement to increase efficiency and value. Monitoring can be used to check capacity and usage.

Simplify retention schedules but remain flexible

Keep retention schedules as simple and broad as possible. It is generally recommended to start with a few simple retention periods and add more as needed. While the majority of record retention can be automated, a modern organization will need to make provisions to accommodate exceptions.

Education and Training

In order to ensure compliance with policy and retention schedules, employees should be educated on the record retention system. Develop training plans and the related tools for educating the employees. Training opportunities should provide periodic refreshers to existing employees and a comprehensive program for new employees.

REFERENCES

[1] Erlanger, L. (2005, June). Information From Start to Finish. InfoWorld, 27(23), 32-36. Retrieved November 11, 2008, from Research Library database. (Document ID: 854090691).

[2] Firms Unprepared for E-Discovery. (2007, November). *Information Management Journal*. Retrieved March 24, 2008, from Business Source Complete database. Available at: https://login.cyrano.ucmo.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=27569851&site=ehost-live

[3] Karp, M. (2006, October). Six steps to ILM implementation. Network World, 23(42), 38,40,42. Retrieved November 12, 2008, from Research Library database. (Document ID: 1158290161).

[4] Power, M & Trope, R. (2006). The 2006 Survey of Legal Developments in Data Management, Privacy, and Information Security: The Continuing Evolution of Data Governance. The Business Lawyer, 62(1), 251-294. Retrieved November 13, 2008, from Research Library database. (Document ID: 1222176271).

[5] Sarbanes-Oxley Act of 2002. (2002). Retrieved on November 13, 2008, from http://fl1.findlaw.com/news.findlaw.com/hdocs/docs/gwbush/sarbanesoxley072302.pdf.

[6] Thompson, Jr., A. A. & Strickland III, A. J. (1995). *Strategic Management: Concepts and Cases* (8th ed.). (pp.276-277). Irwin: Chicago.

[7] Xiaomi, A. (2003). An integrated approach to records management. Information Management Journal, 37(4), 24. Retrieved November 13, 2008, from Research Library database. (Document ID: 378172381).