

ENTERPRISE DATA MANAGEMENT SYSTEMS

**James Garling, Professional Services
EMC Corporation
Edwards AFB, California**

**David Cahill, Product Engineer
EMC Corporation
Hopkinton, Massachusetts**

ABSTRACT

This paper discusses ongoing regulatory effects on efforts aimed at developing data infrastructures that assist test engineers in achieving information superiority and for maintaining their information, and on possible architectural frameworks for resolving the engineer's need versus the regulatory requirements. Since current commercial-off-the-shelf (COTS) Enterprise Content Management (ECM) systems are targeted primarily at business environments such as back office applications, financial sectors, and manufacturing, these COTS systems do not provide sufficient focus for managing the unique aspects of flight test data and associated artifacts (documents, drawings, pre-test data, etc.). This paper presents our ongoing efforts for deploying a storage infrastructure independent enterprise data management system for maintaining vital up-to-date information and for managing the archival of such data.

KEY WORDS

Flight Test, Telemetry, Data Storage, Archival, Regulations, 5015.2, Standards, and Networks.

INTRODUCTION

A study conducted in 2000 reviewed the make up of all the information in the world today. At that time, it was estimated that 75 percent of this content was fixed in nature. Fixed content is defined as records that are typically written once and then never changed. Examples of fixed content records include legal contracts, technical specifications, vendor specifications, purchase orders, wiring diagrams, installation diagrams, quality assurance documents, and flight test data, just to name a few. This content is further characterized by requirements for online access, guaranteed authenticity, and long-term retention. Typically, such records serve as the corporate memory of daily business and flight test program actions or events and enable development and testing organizations to later review, analyze or document the specifics of those actions and events. These fixed content records are, in effect, the evidence of transactions that enable organizations to support current and future management decisions, satisfy testing obligations, achieve regulatory compliance, and, in some cases, protect against adverse litigation.

In a recent study conducted by the University of California, Berkeley, the researchers noted that a reference information “Big Bang” is approaching (with total human information sky rocketing from a few billion gigabytes in 2000 to over 57 billion gigabytes in the coming year). Underlying this research was the question of whether we really need to retain all of this data? The answer is a qualified yes, but without a clear demarcation of what data must be retained versus what can be purged. In a recent poll of information technology (IT) managers, 91% indicated they were storing unneeded data and that deleting at least some of that data was necessary and wouldn’t be noticed by anyone [DMG 2002]. But the reality is that IT managers are not in control of their own destinies when it comes to managing data retention. Further impacting the decision of what can be kept versus what can be thrown away is the plethora of more than 15,000 federal and state laws and regulations in the United States dealing with records retention and how records should be stored, including the requirements that data must be kept in its original form for years, decades, or in some cases forever (such as for some NASA geo-satellite obtained data) [Cahill 2003, DMG 2002].

With this in mind, we see the focus shifting from what must be retained to the process by which records should be maintained. One of the most important components of the record retention process is how the accuracy, reliability, and trustworthiness of the records are maintained over time. This involves not only the protection and security of the media on which the records are stored, but also the integrity of the content. Jolted by recent events in business, finance, and geopolitics, regulators and legislative bodies have acquired a newfound appreciation for the value and importance of records. In addition, the digital age and the increasing prevalence of electronic records have forced these regulators and legislators to revise promulgated legal requirements and standards previously passed that were created when paper, microfilm, and tape were the only viable recording mediums.

For organizations seeking to implement solutions that adhere to these increasing regulatory requirements while maintaining competitiveness in an ever-changing business world, conventional archiving technologies have proven inadequate. Labor-intensive management, data accessibility, and slow retrieval times are just three of the many inadequacies of the fixed content records storage solutions available until now. Recognizing this gap, a new magnetic-disk-based archival technology optimized around the cost-effective retention, protection, and disposition of fixed content records has also been developed.

This paper presents an overview of several of the major regulatory drivers that resulted in the development of this new technology, and then delves into a description of how this technology works and the shortcomings that must be addressed to utilize the commercial solutions in the DoD test industry, while facilitating fulfillment of records-retention compliance. The specific intent of this paper is to provide additional support for software technology refreshes to upgrade centers and data storage facilities that share in the responsibility of processing and managing telemetry information.

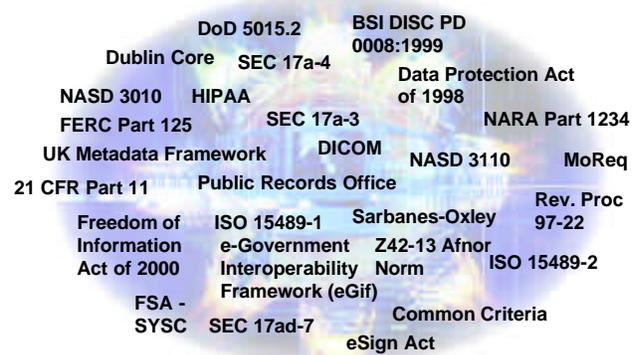
THE REGULATORY WORLD

Every industry has them and internationally, each country has their own for each industry. They are regulations and laws regarding the retention and preservation of electronic records. Whether it is financial services, life sciences, healthcare, legal, manufacturing, government, telecommunications, or utilities, each industry has a different set of standards that companies and organizations within that particular industry must abide by. Stemming from recent fraudulent activity of public

companies, the ever-increasing quantity of electronic records (e.g., in the flight test industry), the explosion of e-mail volumes and its prolific adverse use in litigation, and from more recent homeland security issues, electronic records regulations are increasing in both quantity and legal importance.

To simplify the regulatory landscape, it is best to look at a four-tier classification of regulations:

1. End-User Compliance – Reliance on the end user to ensure adherence to the specified standards (Example: HIPAA, 21 CFR Part 11, Sarbanes Oxley, FERC Part 125)
2. Application Compliance – Reliance on the application provider to have its application certified as compliant with specified standards before it is sold to end users (Example: DoD 5015.2, “Design Criteria Standard For Electronic Records Management Software Applications”)
3. Media Compliance – Specifically targets attributes of the storage media that must be incorporated into an end-user solution in order to be in compliance (Example: SEC Rule 17a-4)
4. Best Practices – These are not required regulations but are instead recommended best practices within a particular industry or environment (Example: DISC 0008:1999 Legal Admissibility & Evidential Weight of Information Stored Electronically, ISO 15489)



Laws and regulations that pertain to records management have a common objective: the creation, retention and preservation, and subsequent accessibility of records that are accurate, reliable, and trustworthy for anyone who may need them throughout their lifecycle. Accordingly, records management laws and regulations typically address what to do and how to do it, with laws focusing primarily on the “what” and regulations dealing more with the “how.” The specificity of the “how” can vary significantly from one regulation to another. Increasingly, both laws and regulations are emphasizing the process by which stipulated objectives are achieved.

Certain laws and regulations focus on specific parts of managing records. For example, there are more than 3,000 that deal with just the storage of electronic records on magnetic media [Cahill 2003]. This component of regulations is concerned with implementing the proper retention and preservation mechanisms to ensure that if and when they are needed, all regulated records can be guaranteed authentic and retrieved in a timely manner. Addressing other records storage issues are components of laws and regulations that deal with policies and procedures, security measures, access requirements, information verification, records migration, and audit trails. In addition, there are a large number of other records management laws that focus on retention policies.

Understanding the laws and regulations that are relevant to each user’s specific business interests – including their legal and regulatory requirements – is very important, but not easy to accomplish.

Like the proverbial iceberg, there is much more that needs to be taken into account than is initially perceived.

To successfully address the issues, problems, and challenges of managing fixed content electronic records requires thinking “outside the box” by applying new perspectives, developing new concepts, and applying new technologies. The storage infrastructure behind this growth must be able to scale accordingly. This environment must be able to consume terabytes of information with minimal storage management overhead, as organizations are reluctant to dedicate significant financial and human resources to the management of these records. It must be able to support unpredictable access to record stores. It is impossible to predict which record will be needed when, but it is a certainty that such a record will need to be retrieved quickly when needed. Think of a patient’s medical record in an acute care situation, or a flight test outcome resting on the production of a critical document. Not having a record in time could have significant ramifications. The subject of medical errors alone has gained recent national attention due to a study by the Institute of Medicine (IOM) published in 2002. The IOM found that medical errors kill an estimated 44,000 to 98,000 Americans every year [IOM 2002].

Add to these requirements the regulatory demands of content authenticity, long-term preservation, and the stringent controls in place for record retention and disposition and you begin to understand the challenges organizations are faced with on a daily basis. With limited financial and human resources, an organization must employ a solution that can cost effectively meet these requirements while maintaining the flexibility to evolve with changing regulations and technology standards.

THE PROBLEMS TODAY

When choosing a solution for managing and archiving regulated records or sensitive / required content there are six key questions that must be answered:

1. Can I verify that my test program and other records will be stored and remain authentic over its required retention time?
2. How will this solution overcome technology obsolescence and ensure long-term availability of my records?
3. Does the solution meet my program and regulatory requirements while adequately supporting the needs of my organization, both today and as it grows?
4. Can this technology be leveraged across the enterprise with other content-generating applications?
5. How will this solution address the data retention and discovery challenges that I am faced with today?
6. Am I prepared to answer a test program’s objectives within stipulated timeframes?

These questions can be answered only through the combination of hardware and software solutions. On the hardware front, the future IT infrastructure will be the heart of successful operations, providing continuous availability of networks (e.g., via Storage Attached Networks a.k.a. SAN, and/or via Network Attached Storage a.k.a. NAS), applications, servers and information storage. In the area of compliant information storage, our approach incorporates a Content Addressed Storage (CAS) solution. This design was targeted at meeting the unique requirements of fixed-content

management, comprised of any form of digitized information asset retained for reference and value, including documents, e-mail, telemetry, satellite photos and imagery, streaming video/audio, X-rays, final form CAD/CAM drawings, etc. Federal civilian agencies and military services that require WORM (write once, read many) storage capabilities will find our infrastructure approach a viable option.

While the hardware layer issues can now be addressed by available commercial technologies, viable Enterprise Content Management (ECM) software that is fully compatible with the DoD flight test industry remains problematic. “Out-of-the-box” ECM products are focused on the commercial world and the types of data encountered there. Although commercial-off-the-shelf (COTS) solutions are capable of being implemented in ways to support most of the current Defense Acquisition Management Framework, these COTS solutions neither encompass integrated support for the variety of flight test data formats, nor do they encompass integrated support for the interface control document (ICD) formats in use in our industry (e.g., TMATS, iDx, and the numerous vendor unique data dictionaries).

In addition to the challenges posed by attempting to incorporate support for these unique formats and data dictionaries, DoD organizations seeking to deploy an ECM also have the added weight of the 5015.2 Records Management Application (RMA) directive. This directive sets mandatory baseline requirements for all RMA software utilized by *any* DoD organization. An important facet to understanding this directive is that it focuses solely on defining what the RMA must do. For example, 5015.2 is based on National Archives & Records Administration Regulations and defines system interfaces and search criteria that must be supported by RMAs that will be utilized by DoD components. To date, only a few COTS products have achieved 5015.2 certification, with the majority having done so by having a CAS hardware architecture underlying their software. Even when certified, COTS products do not contain the necessary components for effectively managing test data in the flight test environment. In addition to the ability to guarantee authenticity and retention of data, RMAs for the flight test industry must have the ability to encompass whole program’s document sets including

- Test Plans;
- ICDs / Metadata;
- Test Data;
- Analysis Algorithms;
- Test Results;
- Test Reports;

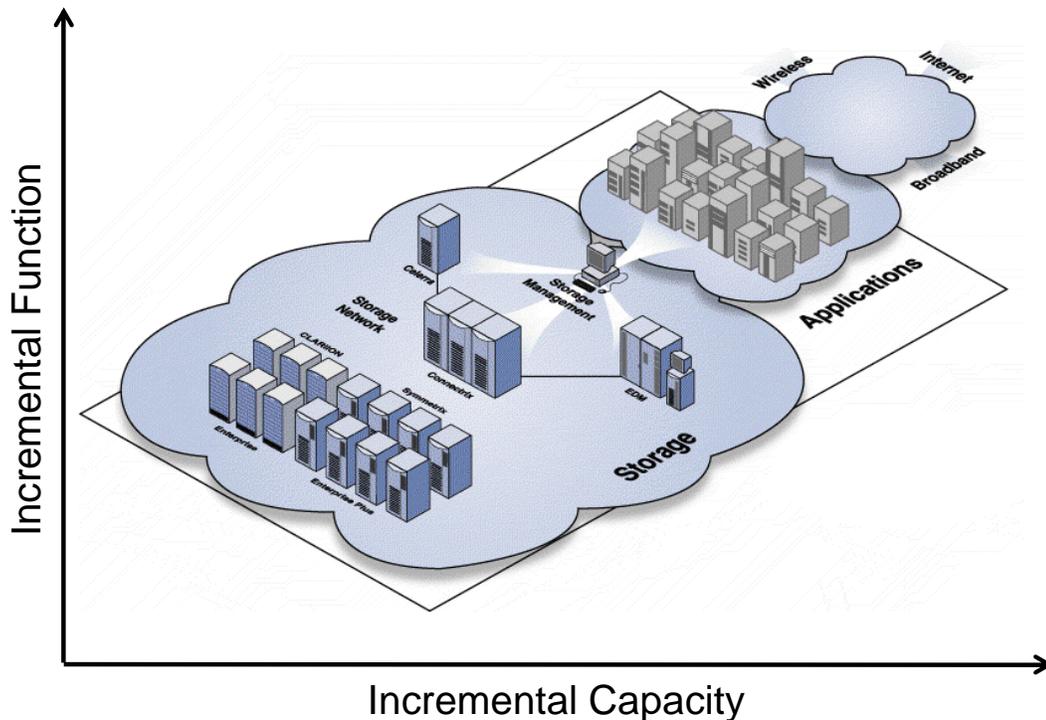
and to encompass logical management and intelligent access to any data type encountered in a test program, including

- Video Sources (Gun, Missile, IR, Weapons Separation, etc.);
- Audio Sources (Cockpit, Air To Air, Air Traffic Control, etc.);
- Bus Data (1553, ARINC 429, H009, Ch 8, etc.);
- PCM Data (Primary, Secondary, Embedded, Ch 4, etc.);
- Combined Data (Ch 10); and,
- *Data formats not yet defined!*

Our efforts have been targeted at resolving this issue by implementing a Repeatable Program Architecture (RPA) that contains components for closing the gap between compliance requirements and actual operational use needs. To this end, we extended our Auto Information Systems (AutoIS) strategy to include a fourth initiative:

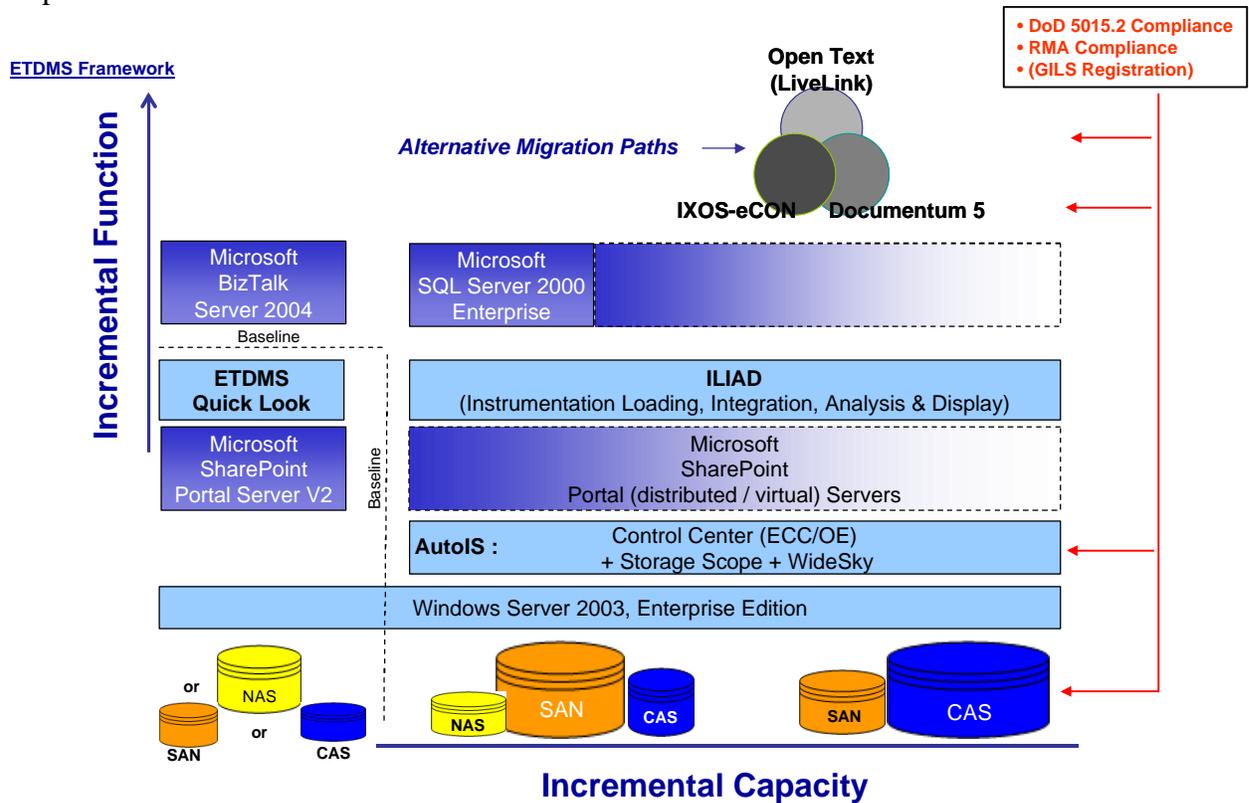
- *Intelligent Supervision* – automate monitoring, reporting, and control of the networked storage infrastructure;
- *Information Safety* – provide intelligent data protection, ensuring the operational integrity of the business;
- *Infrastructure Services* – provide a logical presentation and seamless mobility of information; and,
- **NEW: ETDMS** – provide intelligent object storage based on groupings of associated data and object management based on access management and rights.

Our goal was to achieve the fourth initiative without relying on a single COTS product platform. The end result would be a repeatable solution architecture that could be implemented using any 5015.2 certified RMA. Realizing that some environments would not be held to a 5015.2 level of compliance, our solution focused on deploying a solution framework that was scalable and incrementally capable of compliance through the use of COTS products. The framework we devised allows an environment to be incrementally increased in functionality, capacity, or both, as mission needs expand or grow more complex.



Our initial architecture was implemented as a 100% web-based deployment, thereby allowing disparate operating system environments to access and benefit from the RMA. The diagram below

depicts an architecture we deployed as our baseline, using the Microsoft portal and back office service engines (see area below and to the left of the dashed gray line). This design was implemented to allow an organization to begin with a much smaller baseline and, as needs grow, to extend into enterprise class solutions.



To address the gap in managing flight test industry data, we implemented a “web part” within the document management framework Microsoft SharePoint, which provided us a front-end to a database repository that tracked the association of test data to all related records. We chose SharePoint as the entry point platform based on cost versus performance considerations.

Where we had IRIG 106 Chapter 10 data formats in use, we noted that the data was self-describing with regards to parameter and stream content (via the TMATS header file standard to a Chapter 10 data file), and simply augmented our object encapsulation by adding an extended XML description that emulated the concept of test cards (also known as “flight cards” by test pilots). This design was analogous to using the “Properties” feature available with any Microsoft Office document or artifact, and provided a method whereby a user could perform data lookups based on the characteristics and parameters used in each data file.

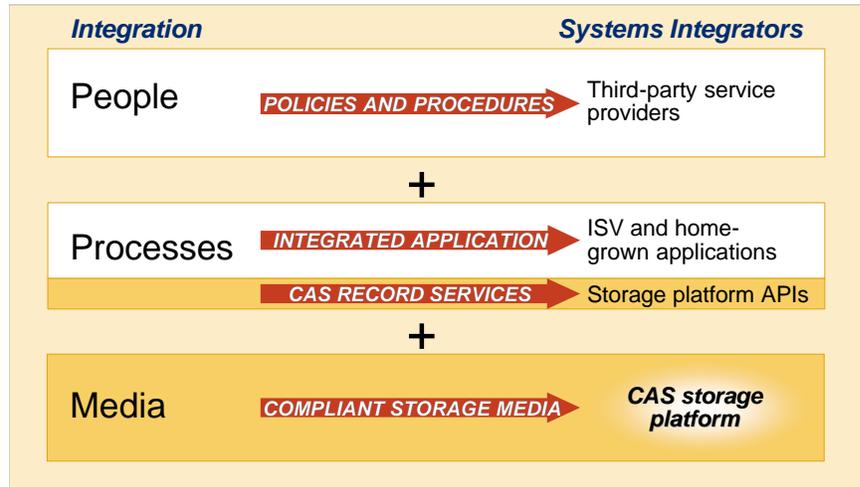
CONCLUSIONS

Regulations and data retention requirements are increasing. At the same time, the volume of electronic records is growing very rapidly, in some instances, at more than 100 percent per year. The need to address these challenges has taken what was traditionally a back-room process and made it a

Program Office and boardroom issue. Ineffective retention and disposition of regulated or legally discoverable information can result in direct fines for regulatory violations, organizational risk due to the exposure of un-disposed records, and the significant cost of maintaining an outmoded technology infrastructure. The policies, processes, and systems you incorporate to manage your sensitive electronic records are a strategic imperative that will have implications on your organization for years to come.

Properly implemented, an ETDMS architecture can provide fast and easy online access to petabytes of information, all the while providing a technology basis for overcoming the limitations and problems discussed here. An ETDMS solution

framework should be capable of solving your most pressing current challenges while providing the power and flexibility to expand across multiple projects and forward into the future. ETDMS based systems are optimized to tactically solve immediate, guaranteed data retention and IT dilemmas, while strategically addressing the need for long-term records management solutions.



BIBLIOGRAPHY

[CAHILL 2003] Cahill, David, "Centera Compliance Edition: Understanding and Addressing the Challenges of Compliance and Discovery," EMC Corporation White Paper; April 2003.

[CRENWELGE 2002] Crenwelge, Robert, and Conway, Brian, and Dillon, Kevin, "Enterprise Flight Data Management System (EFDMS) and Storage Infrastructure Technology Discussion," in Proceedings of the International Telemetry Conference / USA 2002, San Diego, California; October 2002.

[DMG 2002] Data Mobility Group LLC, "Distilling Out Fixed-Content Data," Data Mobility Group White Paper; November 2002.

[IOM 2002] The National Academies - Institute Of Medicine, Annual Report to Congress; 2002.

[RIVEST 1992] Rivest, Ronald L., PhD, "Networking Working Group Request For Comments: 1321," MIT Laboratory for Computer Science and RSA Data Security, Inc; April 1992.

[SEC 2002] Securities and Exchange Commission, Annual Report; 2002.