

## **Solutions for the Automotive Industry in Meeting the Legislated Demands of Record Retention and Accessibility**

*A Frost & Sullivan White Paper Sponsored by EMC  
Analyst and Author: Jarad Carleton*

*In the world of manufacturing, the automotive industry stands out as a highly visible global business with operations on every populated continent in the world. It is also an industry that has been subject to heightened levels of regulatory control combined with the threat of tangible financial loss when fines are assessed for regulatory noncompliance. When manufacturers are brought to court and/or fined for noncompliance, tangible financial losses are easy to measure. However, intangible financial losses will continue for months or years as a bad corporate image pervades the very market the manufacturer wishes to target. Participants in the automotive industry will recognize that similar situations have occurred within the industry in the past. What is important to know, however, is that future occurrences of similar scenarios can be minimized or avoided entirely with the application of proper technology.*

*This white paper describes the data retention issues facing the global automotive industry. As new laws are enacted to regulate discrete aspects of the automotive industry, manufacturers are responsible for maintaining product data for several years to decades as well as ensuring that the retained data is readily available on demand around the clock. EMC Corporation's Content Addressed Storage (CAS) solution, Centera, is examined in the context of the global automotive industry and how it can cost-effectively help meet legislated data retention and availability requirements for both OEMs and suppliers. The introduction of Centera from EMC to automotive manufacturers represents a technological and cost-effective leap forward for long-term storage of fixed content information within a regulatory environment. Fixed content is unchanging digital information assets retained for active reference and long-term business value.*

The information presented in this publication is based primarily on interviews and therefore is subject to fluctuation. Frost & Sullivan takes no responsibility for any incorrect information supplied to us by manufacturers or end-users.

This publication may not be downloaded, displayed, printed or reproduced other than for non-commercial individual reference or private use within your organization, and thereafter it may not be recopied, reproduced or otherwise redistributed. All copyright and other proprietary notices must be retained. No license to publish, communicate, modify, commercialize or alter this document is granted. For reproduction or use of this publication beyond this limited license, permission must be sought from the publisher.

For information regarding permission, write:

Frost & Sullivan  
1040 East Brokaw Road  
San Jose, CA 95131-2309, USA

## Record Retention Trends, & Opportunities

Record retention has always been necessary for manufacturers, however as laws change, demands placed upon record retention have increased in scope and decreased in average time allowed to access data. This data takes multiple forms, but can be segmented into two categories:

1. Dynamic data – information that is actively being edited and changed.
2. Fixed content – unchanging digital information assets retained for active reference and long-term business value.

This latter form represents more than 75 percent of information retained by manufacturers.

In the 1990s fixed content was often stored in a variety of content repositories, a small amount on disk-based storage and the majority on offline optical disk or magnetic tape. The result was often several unrelated storage systems accompanied with the challenges of maintaining them. The problem reached its peak when disparate enterprise content management (ECM) systems emerged to manage different types of fixed content ranging from documents to Web pages and digital media. Having different content repositories, for each type of ECM system placed undue strain on IT personnel and budgets and made the shortcomings of multiple storage repositories painfully obvious.

The opportunity for the storage industry was to develop a cost-effective, online, networked repository for use in the enterprise environment by applications and databases alike. The requirements for a fixed content storage solution are much different than for traditional transaction-based storage:

1. Access times need to be much faster than tape or optical solutions if fixed content is to be readily available to users, yet access time doesn't need to reach the speeds required to support engineers designing products. Therefore, an optimum storage system could leverage the latent knowledge of Internet Protocol (IP), which is cost-effective, and that all IT departments possess in order to network the storage system to business-critical applications.
2. Since the fixed content is retained for long periods of time, the storage solution must:
  - a. Ensure content authenticity.
  - b. Require minimum human intervention.
  - c. Take advantage of technology evolution without requiring information migration.

## The TREAD Act Raises the Specter of Non-Compliance for the Automotive Industry

With the enactment of the TREAD Act (Transportation Recall, Enhancement, Accountability & Documentation) by the U.S. Congress, in November of 2000, numerous issues have surfaced that must be addressed by all automotive industry market participants. Although the U.S. Congress provided the automotive industry with a grace period in excess of two years to

prepare for TREAD compliance, that period came to a close on April 1, 2003 when the act was passed into law. Because the TREAD act calls for nine years of data retention, the amount of data that must be tracked, retained, and offered for retrieval on demand is staggering and creates new IT problems that market participants have not had to cope with before.

In order to comply, all automotive manufacturers in the U.S. and those importing vehicles for sale within the U.S. must track and maintain data on:

- Field reports related to problems with components and systems in the U.S.
- Production data on the number, type, and model of vehicles manufactured each quarter and for the previous nine years. This data must also include manufacturing data on tires and other critical automotive components and systems.
- All documentation sent to or made available online to multiple dealers, distributors, owners, lessor, or lessees in the U.S. that addresses recalls, advisories, manufacturer initiated repair or replacement of the automobile or critical component parts, and customer satisfaction campaigns
- Accident fatalities that have resulted in claims against the automotive manufacturer in the U.S. and abroad if the vehicle is similar to one sold in the U.S.
- Automotive accidents that have resulted in claims against the manufacturer in the United States.
- Claims of property damage attributed to suspected automotive defects
- Consumer complaints that have been filed with the manufacturer regarding systems or components in or on the automobile
- All warranty claims information as it relates to problems with critical safety-related systems or components on an automobile for sale in the U.S.

In addition to the new problems relating to data retention, reporting, and rapid data availability, the U.S. government gave enforcement authorities the ability to not only fine companies for non-compliance in the amount of \$5,000 per day to a maximum of \$15 million for a related series of daily violations, but also included the possibility of criminal penalties against companies and individuals in extreme circumstances. The main compliance concern for manufacturers is the ability to promptly provide pertinent data currently stored on tape, optical, hard copy, or from a handful of IT systems that don't interoperate.

Automotive manufacturers and suppliers do retain the needed information, however the fear of incurring financial penalties stems from an inability to report pertinent information in a timely fashion to the government or the public. Just as each type of information is created in different IT systems, it is also stored in disparate ways over the useful life of the data. Therein creating the difficulty for manufacturers to comply with TREAD.

Because TREAD dictates that all data be available on demand, retention of TREAD data in hard copy form or on tape storage media is not a viable option for compliance. Immediate data availability is critical and TREAD data stored on tape or hard copy can take hours to days to access. Conversely, maintaining fixed content, such as TREAD data, on SAN or NAS-based storage systems, which are optimal for maintaining dynamically changing data in

the engineering environment and where data is retrieved virtually instantaneously, is most likely not a cost-effective solution. The best solution should understand the manufacturer need for a reliable, online, fault-tolerant, and cost-effective storage solution that addresses fixed content TREAD, European End of Life Vehicle (ELV), and e-mail data that ensures the information has not been altered.

### **EU Directive on End of Life Vehicles (ELV) Raises Additional Compliance Issues**

Although TREAD is a significant regulatory factor in the U.S., it is not the only issue automotive manufacturers have to deal with in a global market.

The European Union has issued a directive, referred to as the ELV, that has the goal of reducing landfill waste and ensuring proper hazardous materials disposal when vehicles reach the end of their useful life. ELV dictates that all vehicles and components should be manufactured with materials that allow for recycling and reuse of a minimum of 85 percent of materials by vehicle weight by 2006 and 95 percent by 2015. ELV also requires manufacturers to track the amount of recycled materials that are incorporated into new vehicles for sale in the EU and provide this data to prospective buyers of vehicles and include it in promotional marketing literature. Dismantling information for the vehicle, including data on hazardous materials contained in the vehicle, must be provided to authorized treatment facilities upon disposition.

Although a different regulatory problem than TREAD, ELV has similar data-retention implications for manufacturers. Terabytes of data must be retained over a multi-year period and are required to be immediately available on demand 24 hours a day. Furthermore, ELV forces automotive manufacturers to have unprecedented visibility into their OE parts suppliers manufacturing processes.

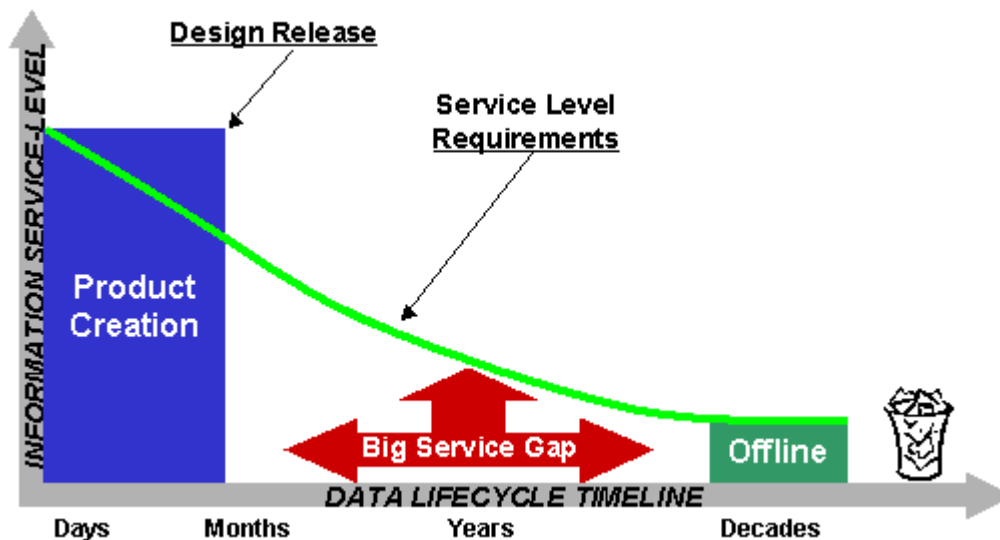
### **E-mail Places Further Requirements on Data Retention and Accessibility**

TREAD and ELV also have implications regarding e-mail retention that require the manufacturer to archive e-mail within the organization if it addresses TREAD or ELV issues. That isn't the only situation in which e-mail retention is an issue for large, publicly owned automotive manufacturers however. Regulatory agencies that govern the trading of securities also require e-mail retention. In these cases, governments have mandated a storage technology that permits information to be stored and prohibits the alteration of that data, the same technology needed by TREAD and ELV.

Because e-mail servers have limited retention capabilities, e-mail must be transferred off the server to a repository. Just as important as retaining terabytes of e-mail to comply with government regulations however, is the need for a managed system that deletes messages automatically and immediately once the retention period is complete. This frees valuable storage space for the manufacturer as well as reduces potential liability by having a consistent and reliably applied e-mail retention and destruction policy.

## Data Storage & the Information Lifecycle in Product Design and Development

The information lifecycle at many automotive companies involves keeping data in the product creation phase online with high performance storage solutions. Immediately after the product design is released however, data is typically stored offline on optical disk, magnetic tape, and in some instances in hard copy form in offsite warehouses. The problem with this model of data storage is that service level requirements for the data in question don't immediately drop to levels low enough to justify offline storage. Components manufacturers, legal teams, dealerships and other groups with a need to access automotive data are subject to delays in data access until the optical disk or magnetic tape can be taken out of storage and be temporarily put back online when requested. The end effect is the creation of a large service level gap that manufacturers aren't able to adequately address in a cost-effective manner.



As networked storage solutions become ubiquitous among automotive manufacturers, the information lifecycle for product design and development will address the changing needs of the manufacturing industry.

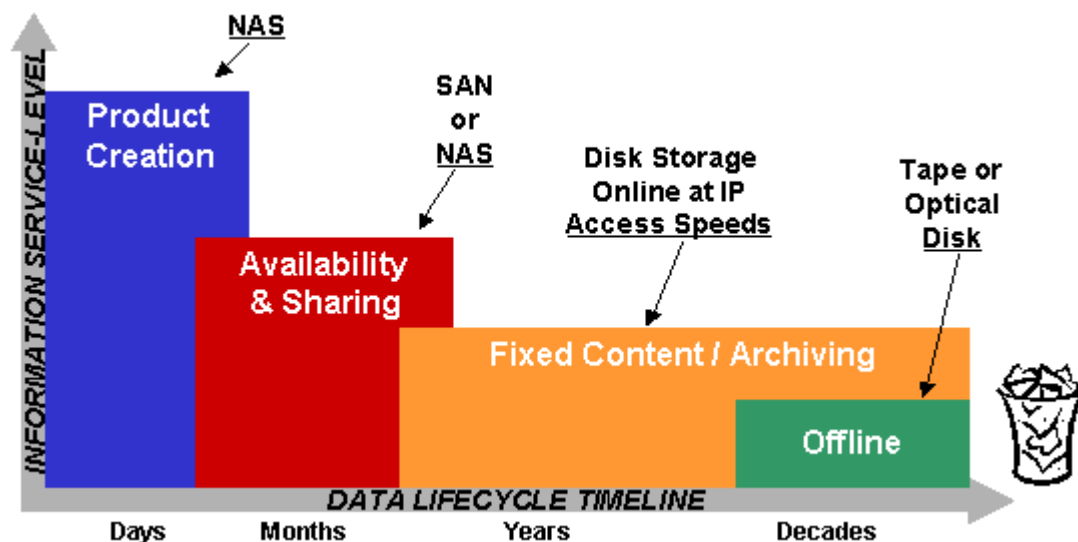
Today, the data lifecycle timeline has evolved into four phases for each vehicle and component part design. Each of the four phases has unique service level parameters, which diminish over time as data moves from one phase to the next. Information lifecycle phases in a modern manufacturing environment are:

- *Product Creation* – Defined typically as a period of days or weeks wherein CAD/CAM work-in-process as well as test and design simulations are conducted. Information in this phase of the information lifecycle is highly dynamic and can change on an hourly basis.
- *Availability & Sharing* – Defined typically as a period of months at which time the product design has been released to manufacturing plants and suppliers. Information is

moderately dynamic in this phase as it is shared with additional design teams and suppliers for reuse or to conduct further engineering tests resulting in altered data. A concern and often the subject of horror stories is the sharing of design documents whose versions were in question. For this reason data leaving the product creation phase at design release may be treated immediately as “fixed content” where information integrity and authenticity for these design files can be strictly maintained, avoiding costly revision control mistakes.

- *Fixed Content/Archiving* – Here the information is no longer subject to changes and becomes fixed content. Multiple users accessing the same fixed content characterize this phase. Historical designs, ECOs, warranty reports, customer complaints, and e-mails (retained for legal purposes) regarding vehicles remaining in their useful life, make up much of this information. Note: If any change needs to be made to a document stored in this phase, the optimum storage solution should keep the original document and create a new version with a unique identifier, such as the example described in the previous paragraph.
- *Offline* – Here information is accessed infrequently, if at all and regulatory requirements no longer encourage manufacturers to keep the information readily available.

As stated earlier, a different type of storage system is needed for the lifecycle of information. To date, a manufacturer had two options characterized as “fast and expensive” or “cheap and slow.” If the storage industry could provide a cost-effective, online solution that ensured information authenticity, then many of the challenges that storage technology placed on IT in the past could be eliminated. In this scenario, all fixed content could be put online, help increase end-user service levels, and be repurposed for other valuable business uses. This is the promise of the newest category of storage, Content Addressed Storage, for fixed content.



## Content Addressed Storage (CAS) and EMC Centera

Content Addressed Storage addresses storage based on a Content Address (CA) not the information's physical or logical placement in the storage solution. EMC Centera™ is the industry's first implementation of Content Addressed Storage (CAS). Centera stores information objects based on a 128-bit globally unique address that is derived from the object's binary representation. With a Content Address derived from the content itself, Centera eliminates the storage of multiple copies of identical information, regardless of how many requests to store a piece of content are made. For business continuance purposes, Centera stores the content and protects it using content mirroring or content parity protection within the same Centera array. The addressing and encryption functions are similar to a public key infrastructure (PKI) ensuring, security, authenticity and nonrepudiation.

When content is stored in Centera, the unique content address for the stored object and the metadata describing the object are inserted into an XML file. The XML file is referred to as a C-Clip™ Descriptor File (CDF), which also has a unique content address calculated for it. The CDF is then protected (in the same way as the object itself) and is the mechanism used by an application to retrieve an information object. Retrieval of content within Centera is based entirely upon the content addresses rather than through the use of a centralized directory, pathnames, or URLs. Using a content address to access fixed content makes the management of physical and logical location of the information unnecessary, which results in a dramatic reduction in system/storage management. In the event that fixed content is altered and stored again, Centera computes a different content address (because the content is different) and stores it in the array. Original fixed content is not overwritten, ensuring an intact audit trail and assurance that fixed content remains in its original state.

As an integral part of maintaining data integrity and audit trail in the event of a hardware failure in one part of the array, Centera will self-heal by detecting the fault and generating a new copy of the content objects. As this process takes place, the affected disk drive or storage node is isolated from the rest of the system until it can be replaced. Lastly, due to the fact that applications don't have knowledge of the physical placement of fixed content within Centera, components can be replaced and Centera software upgraded without disruption, demonstrating a solution architected to easily scale up to one petabyte and beyond.

Centera can and has been configured to help meet the most stringent requirements of regulated environments. Specifically, Centera enforces application-based retention periods within its microcode. Manufacturers have the ability to lengthen the retention periods, but cannot shorten them.

Just as important as data retention in a regulatory environment however, is the importance of automatically deleting unwanted information using U.S. Department of Defense data destruction standards, overwriting information multiple times with random characters, complimentary values, ones, and zeros, when data reaches its expiration date. This not only frees IT personnel from doing a low level maintenance task, but also creates a standard data destruction policy within an organization eliminating potential legal liabilities, and negates any ability to recapture deleted information using disk management tools. Also essential to

any data management policy, is the ability to immediately suspend data destruction in event of litigation. Centera's central policy management allows immediate extension of data retention periods as needed for any eventuality.

### **EMC Storage Solutions Potential for Automotive Industry Record Handling**

As evidenced earlier, TREAD and ELV require data reliability, authentication, access on demand, cost effectiveness, and centralized control. EMC Centera addresses each critical point, directly responding to the needs of the automotive industry.

Data reliability within Centera is assured as a result of its RAIN architecture that eliminates all single points of failure within the platform and enables non-disruptive servicing of the system. Centera itself is composed of independent nodes with one terabyte of raw storage capacity, and is interconnected to all other nodes in the cluster via CentraStar™ software (Centera's operating environment) and a private LAN.

Authentication of content is handled by Centera's use of content addressing, which stores fixed content within the cluster using a globally unique address derived from an object's binary representation that is subsequently time/date stamped. The system operates similar to a public key infrastructure (PKI) that establishes encryption algorithms ensuring the integrity of the data so that the end user knows it has not been tampered with.

Centera easily handles access to fixed content on demand via LAN connectivity to application servers. Because each node within the cluster operates as a storage node or an access node, performance can be scaled to meet demand by non-disruptively adding additional front-end access nodes to augment bandwidth to application servers. This enables Centera to work within a growing database-driven environment that is common in the automotive industry. Centera makes this possible by allowing database fields to interact with Centera's API to a content address as a pointer to specific objects in the cluster. Thus, when a database request is placed regarding a component part of a vehicle, the application will use the Centera content address to retrieve the information.

Centera's contribution to modern cost-conscious manufacturing environments is exemplified in two strikingly obvious ways. The first is the manner by which Centera ensures one copy and one replica of fixed content is stored on the system regardless of the number of times it is used. Thus, when Centera is used to store e-mail in a regulatory environment, it will only keep the original and protect that information via content mirroring or content parity protection, even when that document is attached to an e-mail. Furthermore, depending upon the needs of the organization, Centera can replicate its content items to another Centera in a different geographic location if catastrophic natural disasters are a material business concern.

In the automotive industry, for example, the efficient storing, protecting, and replication of information within Centera substantially lowers the TCO for automotive manufacturers.

Centralized control of one to many Centera clusters also plays an integral role in lowering the TCO of long-term online storage. Key to lowering TCO for long-term storage of fixed

content is Centera's ability to provide a single repository for multiple applications. This translates into measurable savings in cost-per-megabyte of storage for IT budgets as maintenance and training costs significantly decrease. Elimination of multiple repositories also dramatically decreases potential mistakes that lead to financial and criminal liability during litigation. In these situations, when an IT team has to manage several repositories, data could be unintentionally destroyed before a technician is able to modify the retention parameters on one or more repositories. Manufacturing environments utilizing Centera are able to extend retention periods for specific information objects or categories of objects as controlled by the users application, effectively minimizing outside liability potential.

## **Conclusion**

Content Addressed Storage, as implemented in EMC Centera, offers automotive manufacturers an organization-wide, managed storage solution for fixed content originating from and accessed by multiple enterprise-class business applications. Centera's ability to meet stringent regulatory requirements can enable manufacturers to avoid fines related to improper data retention and subsequent intangible financial losses from negative perceptions of the manufacturer resulting from publicity surrounding regulatory noncompliance. In addition, it provides a cost-effective networked storage solution to address the need to provide fixed content to governmental bodies, customers, and dismantling facilities immediately on demand, 24 hours a day. These points taken in conjunction with Centera's ability to eliminate multiple application repositories, thereby reduce IT management and maintenance costs. This increases long-term storage system ROI and makes Centera an optimal solution for fixed content information in automotive manufacturing environments.

## **About Frost & Sullivan**

Based in San Jose, California, Frost & Sullivan is a global leader in strategic growth consulting. This white paper is part of Frost & Sullivan's ongoing strategic research into the Information Technology industries. Frost & Sullivan regularly publishes strategic analyses of the major markets for products that encompass storage, management, and security of data. Frost & Sullivan also provides custom growth consulting to a variety of national and international companies.