symantec.

Confidence in the connected world.

Ensuring Object Integrity and Recoverability within Enterprise Content Management Systems

A white paper by Symantec and CYA Technologies

White Paper: Symantec Business Benefits

Ensuring Object Integrity and Recoverability within Enterprise Content Management Systems

Content

Understanding the Components of Enterprise Content Management

Enterprise content management (ECM) systems present a unique data protection challenge within enterprise-wide backup and recovery practices. ECM systems such as EMC Documentum, IBM FileNet P8 and Open Text Livelink are intended to help companies more easily manipulate and leverage vast amounts of complex data types. An ECM system dynamically creates and manages the entire content lifecycle along with its associated metadata, which is comprised of audit trails, approvals, revisions, annotation, alternate formats and custom attributes, among other properties. This complex information record is often referred to as an Object within the ECM community.

ECM systems are uniquely architected to house a content repository with folders and documents, while the related metadata resides separately in a database. The application architecture itself monitors and maintains the relationships between content and its metadata to ensure overall data integrity, but only as long as the application itself remains intact and synchronized.

The Importance of Metadata

Metadata is defined as "data about data." One of the core values of an ECM system is its ability to maintain extensive amounts of complex metadata about the various documents it manages - data that can be just as crucial as the documents themselves. This is especially true as companies strive to comply with the various regulations that are specific to the preservation and recoverability of authentic metadata, including but not exclusive to the Sarbanes-Oxley (SOX) Act of 2002 for all public companies, the Food and Drug Administration (FDA), Federal Aviation Agency (FAA) and Environment Protection Agency (EPA) regulations, Security and Exchange Commission (SEC) policies for the financial services industry, and the Health Insurance Portability and Accountability Act (HIPAA) for the healthcare industry.

Additionally, within eDiscovery the ability to accurately recover metadata is just as important—if not more so—than the ability to recover business applications and documents. For example, consider the Discovery Rules Amendment of the Federal Rules of Civil Procedure, effective December 1, 2006, which stipulates that companies involved with litigation must, by default, not only produce all requested documents and e-mail in their normally used electronic state regardless of the complexity of the information, but also produce the accompanying

unaltered electronic audit trail. This requirement now eliminates an organization's ability to pay a minimal fine for the failure to produce complex objects within litigation or inspections.

Beyond the compliance and regulatory issues, however, are efficiency requirements for ECM systems. Primary business drivers behind the implementation of these systems are to gain efficiencies within business processes and to improve fiscal performance. Application failures and user mishandling that result in lost metadata will put those efficiency goals in jeopardy. If a company's only option is to roll back its ECM system to the last known good backup, it will lose all the work that employees generated after that point in time. In addition, the company loses significant worker productivity and the ability to process transactions while the ECM system is unavailable. Just in the Life Sciences industry alone, it's anticipated that for every day an ECM application is offline, there's a daily cost to the organization of $1 million due to lost productivity, halted test process cycles and delayed market availability of products(1).

The New Recovery Paradigm: Metadata

A traditional enterprise backup and recovery solution is designed to capture and recover applications at the server level to protect against natural disasters and full system failures. However, compliance and regulatory mandates now require the recovery process to deliver authentic metadata as well as its parent content. With a traditional backup approach, the separate content and metadata servers are susceptible to inconsistencies and corruptions. While the metadata database may be recoverable, the recovered data is essentially unacceptable as the paths or connections between the metadata and the working documents may be lost or corrupted. Recreating those connections becomes an overwhelming time and resource burden or, more likely, is impossible when a system encounters thousands of disconnected or orphaned objects.

Partial Data Loss: The Most Common Data Loss Culprit

Despite the universal commitment to enterprise-wide protection against system failures, the majority of corporate data loss is actually caused by administrative and programmatic errors, user mishandling, accidental deletions, malfeasance, application corruptions, and viruses. These types of incidents are known as partial, operational, or logical data loss.

**Customer Use Case
Cytec Industries
New Orleans, LA**

In their Enterprise Content Management (ECM) center of excellence, Cytec is running Symantec Veritas and CYA Technologies solutions to ensure the most rapid, thorough recovery of system-level and granular ECM information. This combined solution allows them to comply with urgent EPA, ISO and other mandates related to information protection.

In 2007, while running a simple query, an ECM administrator inadvertently forgot the 'where' clause and immediately removed 28,000 items from users' inboxes, halting workflows because links enabling users to take the next step in the process had been removed.

With CYA SmartRecovery, the administrator easily identified the missing links in the CYA Capture Sets and ran a job to restore them back to their original, authentic state without taking the application offline.
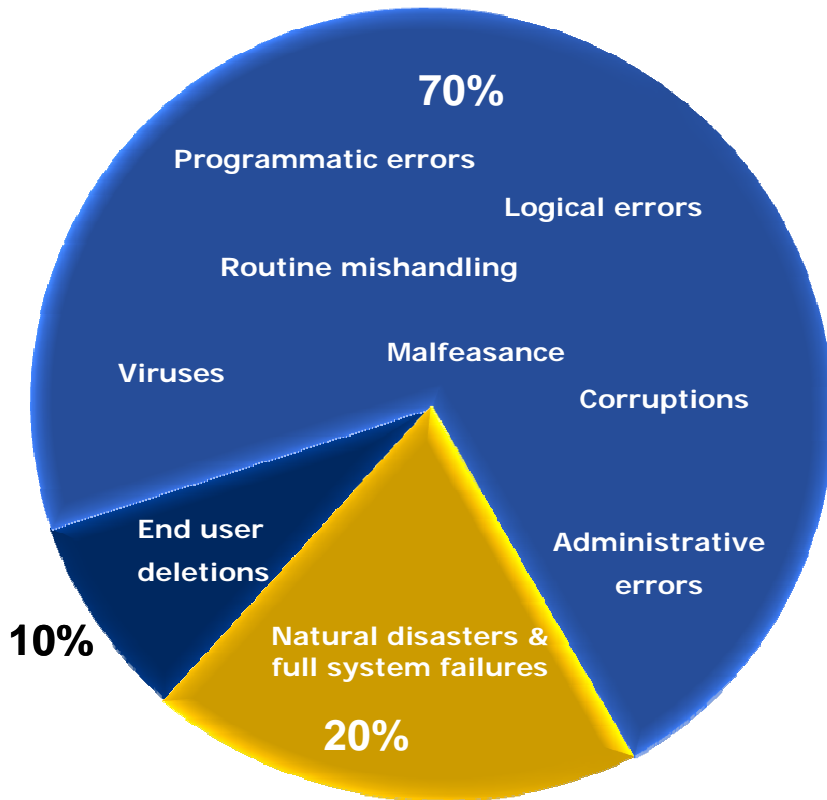
Image 1. Causes of partial (operational) information loss within ECM[2]

Partial data loss is most harmful within ECM when it causes repository inconsistencies that disconnect the complex metadata (workflow states, electronic approvals and audit trails) from their parent content. Although this data loss does not result in a full system failure, it results in extensive disruptions within the business, potentially impacting the majority of users and causing widespread information loss, sometimes in the thousands of documents and their associated metadata. During these incidents, organizations learn quickly that their only options are to recover using their enterprise backup system or conduct a manual recovery of select documents from tape. Either scenario requires excessive IT resources, risks loss of new and updated information generated since the last backup, and introduces potential corruptions between content and its metadata during the recovery process.

(2) AIIM International, Strategic Research Partners, EMC and IBM

5

Historically, once a record or group of records had been lost or corrupted within ECM, the individual record recovery process likely involved pulling one or more backup tapes, combing through the tapes to find the relevant documents, and then recovering the document without metadata outside of the application. If possible, users would attempt to manually rebuild rudimentary audit trails, but only as an afterthought, and with great time and effort. In today's regulatory environment, it is no longer an option to exclude and not validate metadata within a backup and recovery practice.

The following examples are real cases of partial data loss within ECM environments and demonstrate the shortfalls and challenges that are encountered when attempting to recover from partial information loss within ECM repositories.

Real World Examples of Partial Data Loss		
Organization:	**Global Drug Manufacturer (F150)**	**Global Civil Engineering Firm (F100)**
Incident	Single user deletes folder with 15,000 documents linked to over 100,000 objects, affecting hundreds of users and making it impossible to identify every lost object.	Administrator accidentally deletes hundreds of complex subcontractor bids out of production folder instead of development folder during routine maintenance, immediately halting the bidding process.
Manual Recovery	• Required more than 1,000 man hours to manually recover only the documents residing on backup tapes. • A full system restore was not an option because critical information added to the repository since the last cold backup would be overwritten.	• Recovered only the bids from backup tapes outside of the repository, and had to manually reload into repository with new ID, author and date. • Full application restore was not an option because operation was global and required application to remain online.
Impact of Loss	• Manufacturing shutdown due to lack of proper documentation required by FDA. • Significant metadata loss, downtime, and user disruptions; excessive burden to IT organization. • During recovery process, discovered this was one of over forty incidents in last 3 years.	• Took temporary workers several months to manually load contracts and recreate some of their known properties in the repository. • After a year, groundbreaking still delayed due to information loss incident; millions of dollars in lost opportunity. • Extreme negative exposure for the engineering firm and their oil company client

The Value of Granular Recovery for ECM

Symantec Veritas NetBackup is the foundation for an enterprise-wide ECM recovery strategy designed to protect against system failures and/or natural disasters, after which customers need to recover entire servers and databases from the operating system on up. However, as previously discussed, the majority
of corporate information loss is associated with human and programmatic errors, inconsistencies or isolated disruptions. Granular recovery - the ability to easily locate objects (whether one or thousands), validate their accuracy, and recover them directly back to their original state within the repository - is now an essential component for a complete ECM recovery strategy.

The ability to quickly perform granular recovery provides immediate value to companies on multiple fronts:

- Offers rapid access to and recovery of complex objects in response to partial loss or corruptions, ensuring business continuity with no application downtime.

- Reduces IT administrative burdens surrounding object-level data recovery by limiting required recovery resources to just a few minutes of a single administrator's time.

- Facilitates compliance with regulatory mandates from the SEC, EPA, FDA, FAA, the Federal Rules of Civil Procedure for eDiscovery, Sarbanes-Oxley, and many others related to the preservation and production of electronically stored information such as complex metadata.

- Minimizes costs related to penalties and legal fees associated with non-compliance that may be discovered during audits and inspections; mitigates financial risks associated with lost revenue caused by application downtime.

> **"A recovery strategy for ECM should include the ability to recover from incidents of partial or logical information loss. Companies with traditional disaster recovery solutions for ECM often possess a false sense of security, and don't realize that if files are inadvertently deleted or corrupted within their ECM system the solution for recovery is to bring the entire system down and restore from the last full backup"** said Laura DuBois, Research Director for Storage Software at IDC. **"CYA SmartRecovery empowers organizations to recover quickly and easily from partial information loss by locating content and its metadata, restoring them to their original state while the system stays online, and compressing the data loss window."**

CYA SmartRecovery: Ensure Object-Level Integrity and Recoverability for ECM

Effectively safeguarding content and its metadata within an ECM system requires a granular recovery solution that understands the nature of metadata, can maintain the physical relationships between it and its related documents, and validate metadata authenticity. Combined, Symantec Veritas NetBackup and CYA SmartRecovery™ from CYA Technologies enable companies to achieve the most stringent service level agreements associated with Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) for ECM systems.

CYA SmartRecovery's primary benefits for ECM deployments are:

1) Live Capture: CYA SmartRecovery automatically captures incremental changes in the repository as frequently as every 15 minutes without any application or system downtime.
2) Integrity: Using exclusive Intelli-Capture™ technology, CYA SmartRecovery checks to ensure integrity between content and its metadata, and proactively flags inconsistencies for immediate corrective action by the administrator.
3) Recoverability: CYA SmartRecovery recovers objects back to their last "original" state within the ECM repository without any system or application downtime.
4) Reduction of Data Loss Window: CYA SmartRecovery supports Symantec Veritas NetBackup with time-based recovery functionality that allows organizations to recover to a specific point in time, minimizing the data loss window within the repository.

In addition, as companies grapple with new records retention mandates, they are implementing records management tools such as records retention services or expiry dates, that enable them to dictate the retention periods for various kinds of data based on established criteria, sometimes within the audit trail itself. CYA SmartRecovery operates as an integrated extension of an organization's records management system, ensuring that data is only preserved up to the established retention period.

CYA SmartRecovery and Symantec Veritas NetBackup

Symantec Veritas NetBackup and CYA SmartRecovery work synergistically to provide enterprises with a complete and seamless recovery solution for ECM environments. They work together to help companies achieve the most stringent service levels associated with data loss windows and application availability in response to disasters and system failures, as well as to protect companies against partial data loss. CYA SmartRecovery works independently within the ECM application and requires no additional integration with NetBackup. To conserve storage space, organizations can migrate their CYA objects to low cost storage devices.

Posing the Tough Questions

ECM platforms present a promising opportunity - as well as a challenge - within today's business environment. Symantec Veritas NetBackup has partnered with CYA Technologies to help their customers ensure business continuity and achieve recovery goals at the granular level for ECM. This partnership offers the ultimate recovery solution for enterprise content management. To comprehend the added value to your existing enterprise backup practices and ensure that your company is compliant with all of the new regulations (see sidebar), ask yourself the questions below.

- **What is the shortest data loss window I can achieve to meet our corporate or federal service level agreements?**
  - CYA SmartRecovery reduces the data loss window to as little as 15 minutes using its exclusive Intelli-Capture technology to provide incremental captures of all changes within the repositories during scheduled jobs.

- **Do I have repository checks in place to eliminate inconsistencies between content and metadata databases?**
  - CYA SmartRecovery runs more than 350 integrity checks to identify corruption or inconsistencies between content and its metadata as soon as it's changed within the repository. It flags corruptions such as missing records, missing content and invalid relationships between records.

- **Can I provide granular recovery of content and metadata?**
  - No other solution is able to recover granular metadata for ECM. With Intelli-Capture, CYA SmartRecovery captures all aspects of new or updated records within the online

Mandates and Regulations for Metadata Preservation and Recoverability

- Sarbanes-Oxley Section 802
- ISO 15489
- DOD 5015.2
- COOP FPC 65
- SEC 17a
- US Patriot Act
- FDA 21 CFR Part 11 & FDA Current Good Manufacturing Processes
- International Oil Industry Standards & Practices
- Petroleum Regulation of 1969: Section 54, 55 and 56
- Critical Energy Infrastructure Information (CEII)
- US Federal Civil Procedure Rules, Discovery Rules Amendment
- UK Civil Procedure Rules

repositories in a single pass with full-text indexing, then rapidly recovers objects back in to the live repository.

- **Is my metadata authentic and compliance-worthy?**
  - It's crucial to ensure that any recovery solution your organization has in place for ECM is able to fully meet the demands of all regulations affecting your organization. CYA SmartRecovery enables organizations to meet the demand for the capture, validation and recovery of information and its associated metadata that is required by numerous regulations including Sarbanes-Oxley, HIPAA, the Gramm-Leach-Bliley Act, international Basel II banking rules, the USA Patriot Act and the Food and Drug Administration 21 CFR Part 11 guidelines.

Summary

ECM systems such as EMC Documentum, IBM FileNet P8, Microsoft SharePoint and Open Text Livelink help companies more effectively deal with ever-growing amounts of data. But safeguarding such systems and their complex data requires a thorough solution that not only protects the entire system against failures and disasters, but also enables object-level recoverability of the content and metadata in cases of partial data loss. Whether in response to everyday incidents or regulatory and litigation-related requests, companies must be able to recover aged, lost or corrupted data. The combination of Symantec Veritas NetBackup and CYA SmartRecovery provides the ultimate recovery solution for ECM users, empowering them to recover individual documents with their associated metadata in mere minutes with minimal usage of resources and no costly ECM system downtime.

(1)Infonetics Research 2005. Costs of Enterprise Downtime: North American Vertical Markets.

About Symantec

Symantec is a global leader in infrastructure software, enabling businesses and consumers to have confidence in a connected world. The company helps customers protect their infrastructure, information, and interactions by delivering software and services that address risks to security, availability, compliance, and performance. Headquartered in Cupertino, Calif., Symantec has operations in 40 countries. More information is available at www.symantec.com.

For specific country offices and contact numbers, please visit our Web site. For product information in the U.S., call toll-free 1 (800) 745 6054.

Symantec Corporation
World Headquarters
20330 Stevens Creek Boulevard
Cupertino, CA 95014 USA
+1 (408) 517 8000
1 (800) 721 3934

www.symantec.com