

# **Data Classification: The Cornerstone for Successful Information Lifecycle Management**

September 2003

No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of EMC Corporation.

The information contained in this document is subject to change without notice. EMC Corporation assumes no responsibility for any errors that may appear.

All computer software programs, including but not limited to microcode, described in this document are furnished under a license, and may be used or copied only in accordance with the terms of such license. EMC either owns or has the right to license the computer software programs described in this document. EMC Corporation retains all rights, title and interest in the computer software programs.

EMC Corporation makes no warranties, expressed or implied, by operation of law or otherwise, relating to this document, the products or the computer software programs described herein. EMC CORPORATION DISCLAIMS ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. In no event shall EMC Corporation be liable for (a) incidental, indirect, special, or consequential damages or (b) any damages whatsoever resulting from the loss of use, data or profits, arising out of this document, even if advised of the possibility of such damages.

EMC<sup>2</sup>, EMC, Symmetrix, AutoIS, Celerra, CLARiiON, CLARAlert, DG, E-Infostructure, HighRoad, Navisphere, PowerPath, ResourcePak, SnapView/IP, SRDF, VisualSAN, WideSky, The EMC Effect, and where information lives are registered trademarks and EMC Automated Networked Storage, EMC ControlCenter, EMC Developers Program, EMC Enterprise Storage, EMC Enterprise Storage Network, EMC OnCourse, EMC Proven, EMC Snap, Access Logix, AutoAdvice, Automated Resource Manager, AutoSwap, AVALONidm, C-Clip, CacheStorm, Celerra Replicator, Centera, CentraStar, CLARevent, Connectrix, CopyCross, CopyPoint, CrosStor, Direct Matrix, Direct Matrix Architecture, EDM, E-Lab, Engenuity, FarPoint, FLARE, GeoSpan, InfoMover, MirrorView, NetWin, OnAlert, OpenScale, Powerlink, PowerVolume, RepliCare, SafeLine, SAN Architect, SAN Copy, SAN Manager, SDMS, SnapSure, SnapView, StorageScope, SupportMate, SymmAPI, SymmEnabler, Symmetrix DMX, TimeFinder, Universal Data Tone, and VisualSRM are trademarks of EMC Corporation. All other trademarks used herein are the property of their respective owners.

Copyright © 2003 EMC Corporation. All rights reserved.

C1059

# Table of Contents

**Introduction .....3**  
**What is Data Classification? .....3**  
**The Compelling Value of Data Classification .....4**  
**Where (and How) Do You Begin? .....4**  
**Conclusion.....5**

# Introduction

Information is the heart of today's businesses. Client databases, product definitions, inventory databases, e-mail stores, and much more now support the around-the-clock demands of today's economy. As the quantity of data proliferates exponentially, IT organizations must have well-executed plans in place to effectively manage and protect that data, both today and for the long term, to ensure their success and viability.

All information, however, is not equal. Active and online data that has a particularly high value needs to be available at all times for rapid access by multiple organizations and applications. Some data requires 100 percent, instant accessibility around the clock, with no tolerance for downtime. Some data is more valuable to certain organizations than it is to others. And some changes in value over time, while other data simply requires archiving for occasional access or long-term storage.

Understanding the value of your data to the day-to-day operation of your business—and thus how quickly it needs to be accessed and by whom—is fundamental to all elements of an information lifecycle management program, including:

- Designing an effective storage infrastructure strategy
- Optimizing storage management to control costs and complexities associated with information growth and storage utilization
- Consolidating information storage to leverage current investments into a more efficient, more manageable information infrastructure
- Planning for business continuity to effectively manage availability of data under all circumstances

Data classification is the essential first step to undertaking any or all of these components of an information lifecycle management program.

## What is Data Classification?

Data classification is a process that defines the performance and availability characteristics of an organization's different sets of data and recommends an appropriate storage technology that meets the needs of each classification. Data in the enterprise is grouped into logical categories to facilitate a key storage objective. For example:

*To create a disaster recovery strategy or business continuity plan, you want to classify your data by business criticality.* This will allow you to create the appropriate linkage between the value of the data to your business and the data recovery measures required to protect that data.

*To define a chargeback model, you want to classify your data by storage resources consumed or by management costs.* This will ensure that departments are carrying the appropriate storage burden commensurate with their use of storage resources.

*To create a comprehensive storage strategy, you want to classify your data by business priority (e.g., first-to-market, customer advocate).* This will ensure that the data is maintained on the appropriate storage infrastructure commensurate with its priority to the business.

*When designing an information storage consolidation plan, you want to classify your data by physical state and location.* This will allow you to identify and eliminate isolated islands of data and bring the most important data items closer to their end users.

# The Compelling Value of Data Classification

Data classification models can add tremendous value to an organization. Effective data classification is an enabling cornerstone to an enterprise-wide data management optimization strategy.

Properly done, a data classification exercise yields a model that consolidates data with similar data management requirements into logical groupings, eliminating the need to describe data items individually. This model can then be used to help define the case for change or to provide the framework that drives the creation of tiered storage architectures.

As your organization grows, the amount of data you need to manage your customers—internally and externally—grows exponentially. The ability to describe and manage this data as individual data items becomes increasingly complex. Grouping this data into logical categories lessens this complexity and provides your organization with the opportunity to better understand, manage, secure, and utilize your data to successfully drive your business. By aligning the detailed storage services with the value of the data in each group, your organization can utilize a completed data classification model to drive overall data management optimization.

Data classification drives compelling value by:

- Providing a clear picture of the categories that exist in your organization
- Describing the linkages between your core business functions and their associated data
- Enabling you to demonstrate the economic value of data to your business
- Enabling you to design and develop a shared technology architecture for each category of data
- Enabling you to define a suite of storage-related services (e.g., availability, recoverability, manageability, scalability, and chargeability) for each category of data

## Where (and How) Do You Begin?

Effective data classification begins by establishing a goal for what you want to do with that data—for example, create a business continuity plan, establish charge-backs for all or a group of departments, or move or consolidate a data center. This drives how you group and segment your data in order to best achieve that specific goal. The desired end result of this exercise is to create a model that is actionable—meaning that it defines a finite and identifiable set of information that can be acted upon in a way that meets your stated business goal.

Some amount of data classification is intuitive and site-specific, driven out of business practices and environmental requirements. However, achieving the best results requires an *in-depth* understanding of how the organization works, the available technology infrastructure, and the specific information needs for each business practice—both for today and for the future—in order to adequately plan for growth and change.

Maximum benefit is achieved from an organized approach to a data classification effort that also utilizes proven methodology and tools to drive the process and elicit the optimum information. The importance of making careful, well-researched, and highly informed decisions on seemingly obvious points, such as what your storage-related goal should be or how different groups perceive the value of data to that goal, cannot be overemphasized. Each decision drives the next and a misalignment at any step in the process can have a significant impact on schedules, budgets, and the effectiveness of the project.

An effective data classification approach utilizes field-proven planning worksheets and targeted interviews to ensure comprehensive and organization-focused information gathering. Application mapping databases and modeling tools also assist in properly assigning a clear *perspective* to the project, which will guide how the information should be best organized in order to achieve that result. Such tools are also particularly effective for assigning data to *schemes*, which define the overall segmentation for the grouping of data; *classes*, which provide meaningful “containers” for grouping data; and *rules*, which serve as the placement criteria to determine which data elements belong in each class. Tools that automate the data collection and classification process can further facilitate and speed the data classification process.

A data classification exercise is an ideal time to consider thinking out-of-the-box about your storage program objectives instead of limiting yourself to existing enterprise rules. If opportunity exists, organizations should consider making a break from how things have historically been done and choose new—and potentially better—ways for doing things in the future.

## Conclusion

Data classification is one cornerstone to achieving the full value of broader storage-related programs. Including data classification as an integral component of a storage-related program provides the opportunity to gather the information necessary to optimize your storage environment and maximize your return on investment.

Data classification should be considered the critical *initial* step that enables a number of valuable storage management capabilities. Proper data classification establishes service-level requirements for tiers of applications and their data, which is delivered by appropriate classes of technology configurations.

The end objective of data classification is ultimately to provide an actionable grouping of data within an enterprise. The resulting data classification model can then be used to either build a business case around current enterprise storage practices or to facilitate the implementation of storage-related program recommendations.

### Data Classification in Perspective: An Example

- Business goal: Reduce costs through storage management optimization.
- Requirements: List of business functions/departments classified by storage resource usage or management costs.
- Data classification: Classify business functions/departments by resource or management costs in order to devise charge-back models based on a combination of information resources consumed and service costs incurred.
- Extension: These chargeback models can then be used to drive the cost of storage and disk utilization metrics in the enterprise by aligning department storage needs with IT storage costs.