



July 13, 2007

## Why You Should Perform Data Classification

Pierre Dorion

What you will learn: Regardless of how daunting a task it may be, understanding the data we store is the first step in addressing data management issues. This tip outlines some of the main reasons why data classification is essential.

Data classification is an exercise every organization will have to consider at some point. The economics of data growth go far beyond simply adding storage capacity. Getting a good understanding of the type of data being stored is at the root of many IT projects that are causing most companies growing pains and headaches. Data classification can be based on type, application, owner, age, legal aspect, criticality, value, etc., and a growing number of software tools are available to assist with this task. But essentially, data must be siloed so it can be better managed. Data classification must be a collaborative effort between IT and business operations since the business units rarely know what data they need, and IT typically only has limited knowledge of who uses the data and why.

Consider the following:

### **How do you decide what data to retain?**

We have all heard of information lifecycle management (ILM), and opinions differ as to what it is, what it is not, and whose responsibility it is. Regardless of opinions and what it is called, at some point or another data might be disposed of, moved offline (i.e., archived) or migrated to lower cost storage (tiered storage). But, before a decision is made regarding where data will reside next, there must be an understanding of what the data is.

### **What data needs to be backed up?**

You are likely backing up data that has not been modified or even accessed in months -- if not years. This practice unnecessarily uses up time in a shrinking backup window, network bandwidth and capacity on your backup storage infrastructure. How long does it take to complete a full backup of your file servers? How much of that data is actually "production" data?

### **What data should be archived?**

Too often, backup products are used as archival tools simply because a decision cannot be made as to when data should be taken out of production and out of the backup loop. It is not rare to see a financial database backup retained for seven years, only to find out that the same database tables are also part of last night's backup.

## **What data should be restored first for disaster recovery?**

When planning for disaster recovery, it is seldom clear which data must be restored first. Recovery time objectives (RTO) are typically driven by the criticality of a business process or application, but the planning often falls short of clearly identifying the associated data.

## **How is data migrated in tiered storage?**

Regardless of whether storage tiers are implemented based on performance requirements, criticality or functionality, data migration across tiers can only take place once data is classified.

## **What data is subject to regulatory compliance laws?**

Regulatory compliance targeting availability of records has sent many companies running in all directions, because they had never taken the time to examine what data is stored. The answer has unfortunately been to increase capacity until there is a better understanding of what is subject to the rules. This understanding is unlikely until the data is inventoried and classified.

## **How do you develop a chargeback model for storage?**

When the time comes to obtain funding for IT, knowing which departments or functional areas are the biggest storage consumers can help build a business case for IT. Data classification can assist with developing a "chargeback" model for storage.

In the heydays of paper records, there were records managers. These people knew what the records were, where they were stored and when they should be archived and disposed of. Nowadays, records managers have mostly been replaced by data storage administrators -- who cannot be as close to the data as their predecessors could.

For many large organizations, classifying existing data may never be fully addressed due to the massive amounts of records accumulated over the years. For some, the only answer might be to draw a line in the sand and develop data management policies going forward that include categorization as soon as data is generated.

About the author: Pierre Dorion is a certified business continuity professional for Mainland Information Systems Inc.

All Rights Reserved, Copyright 2000 - 2008