The Importance of Data Classification and Ownership

by Carol Woodbury

Because of laws such as the Health Insurance Portability and Accountability Act (HIPAA), the requirements of Sarbanes Oxley (SOX) auditors, and data breaches, organizations are beginning to realize that they must secure their data - that is, object level security must be implemented properly.. Thus, organizations are increasingly classifying their data and identifying its appropriate owners. Proper classification of data is essential to ensuring that data is secured correctly. This article details the factors you will want to consider as you go through the process of classifying your data.

What is data classification?

Data classification entails analyzing the data your organization retains, determining its importance and value, and then assigning it to a category. Data that is considered "top secret" (whether contained in a printed report or stored electronically) needs to be classified. Why? So that it can be handled properly. IT administrators and security administrators can guess how long data should be retained and how it should be secured, but unless the organization has taken the time to classify its data, it may not be secured correctly or retained for the required time period.

When classifying data, determine the following aspects of the policy:

- **Who has access to the data.** Define the roles of people who can access the data. Examples include accounting clerks who are allowed to see all accounts payable and receivable but cannot add new accounts, and all employees who are allowed to see the names of other employees (along with managers' names, and departments, and the names of vendors and contractors working for the company). However, only HR employees and managers can see the related pay grades, home addresses, and phone numbers of the entire staff. And only HR managers can see and update employee information classified as private, including Social Security numbers (SSNs) and insurance information.
- **How the data is secured.** Determine whether the data is generally available or, by default, off limits. In other words, when defining the roles that are allowed to have access, you also need to define the type of access—view only or update capabilities—along with the general access policy for the data. Many companies set access controls to deny database access to everyone except those who are specifically granted permission to view or update the data.

  Note: Notice I have not stated the i5/OS security setting for the file—I have defined the access in general terms just as I described who should have access in general terms. Determining who has access and identifying the i5/OS security settings will come when the data custodian (described later in this article) implements this policy.

- **How long the data is retained.** Many industries require that data be retained for a certain length of time. For example, the finance industry requires a seven-year retention period. Data owners need to know the regulatory requirements for their data, and if requirements do not exist, they should base the retention period on the needs of the business.
- **What method should be used to dispose of the data.** For some data classifications, the method of disposal won't matter. But some data is so sensitive that data owners will want to dispose of printed reports through cross-shredding or another

secure method. Or they may require employees to use a utility to "scrub" their PCs after they erase files containing sensitive data.

- **Whether the data needs to be encrypted.** Data owners will have to decide whether their data needs to be encrypted. They typically set this requirement when they must comply with a law or regulation such as the Payment Card Industry (PCI) Data Security Standard.
- **What use of the data is appropriate.** Before data security became such a hot issue for organizations, people in many roles within and outside the company used data in all types of reports. This aspect of the policy defines whether data is for use within the company, is restricted for use by only selected roles, or can be made public to anyone outside the organization. In addition, some data has legal usage definition (for example, California has defined the appropriate use of a Social Security number). Your organization's policy should spell out any such restrictions or refer to the legal definitions.

Let's face it—security administrators don't have extra time on their hands. Classifying data is beneficial because it helps security administrators and internal auditors focus their attention on the data that is most critical to the business, thus ensuring that it is secured and handled properly. Not that other data is ignored, mind you, but if administrators can check the access controls on only a limited number of databases or applications in a given time period, at least it's clear on which ones they should spend the majority of their time.

Proper data classification also helps your organization comply with pertinent laws and regulations. For example, classifying credit card data as private can help ensure compliance with the PCI Data Security Standard. One of the requirements of this standard is to encrypt credit card information. Data owners who correctly defined the encryption aspect of their organizations' data classification policy will require that the data be encrypted according to the specifications defined in this standard. Classifying data as private can also help your organization comply with the various data breach notification laws that many states have enacted. (The State PIRG Consumer Protection Web site, www.pirg.org/consumer/credit/statelaws.htm#breach, can help you keep track of the states that have enacted the notification laws.)

What classifications should be used?

There are no hard and fast rules about the titles and number of classifications. The general guideline is that the definition of the classification should be clear enough so that it is easy to determine how to classify the data. In other words, there should be little (if any) overlap in the classification definitions. Also, it is helpful to use a term for the title of the classification that indicates the type of data that falls into the particular category.

Here are some examples of categorizing data by title:

- **Private.** Data that is defined as private, such as SSNs, bank accounts, or credit card information.
- **Company restricted.** Data that is restricted to a subset of employees.
- **Company confidential.** Data that can be viewed by all employees but is not for general use.
- **Public.** Data that can be viewed or used by employees or the general public.

Data classifications can also change. For example, IBM will often classify new i5/OS release information as IBM Confidential Until Announced. The recipients of this information can properly protect and use the information before the announcement and can then more freely use the information after IBM formally announces a new release.

What are the "right" classifications?

There is no right or wrong classification of data. Remember, data classification is supposed to ensure that business assets are properly handled. If your organization's management does not care about its vital business asset—data—all of the data can remain unclassified. If the data is lost or stolen or otherwise inappropriately used, there is no one to blame but the management personnel who decided not to classify the data.

However, I encourage you to at least identify and classify any private information that your organization retains. Also, classify all the data that is vital to your business. Data such as a retailer's vendor lists, a transportation company's pricing information, a medical device company's product specifications, or any information that could be used by a competitor to harm your business should be classified to ensure that the data custodian secures it properly.

Who decides data's classification?

The individual who owns the data should decide the classification under which the data falls. The committee that wrote the data classification definitions or policies can certainly help or provide guidance, but the final determination for the classification should be the data owner's responsibility. The data owner is best qualified to make this decision because he or she has the most knowledge about the use of the data and its value to the organization.

The database administrator (DBA) can be a good checkpoint to ensure that data is classified and protected properly. Data owners set the classification, but the classification may be poorly communicated or forgotten by programmers developing in-house written applications. When new files are created, the DBA can review the classification to ensure that programmers understand the type of data with which they're working. When new files are moved from the development environment to production, DBAs can perform a final check to ensure the default access on the file is being set appropriately, given the data's classification.

Finally, data owners should review their data's classification at least annually to ensure that the data remains correctly classified. For example, if data owners had been reviewing data classifications for the past few years, they probably moved much of their employees' information—especially information such as SSNs—from a "confidential" classification to a "private" classification. SSNs were never considered private until they were used for identity theft. Since thieves started to steal databases of SSNs, their classification has been upgraded to restrict access and more tightly control their use.

Will the real data owner please stand up?

In addition to classifying data, an organization needs to assign an owner. The owner is not the i5/OS or OS/400 user profile that owns the database object on the system; rather, it is the person in the organization who owns the data that is stored in the database on the iSeries. The data owner is typically a director, or at least a department head, who has a vested interest in making sure the data is accurately and appropriately secured. Take a financial application, for example. Depending on the size of the organization, the data owner may be the CFO or one of the directors who reports to the CFO. The person who is appointed needs to understand the importance and value of the information to the business as well as the ramifications of inaccurate storage or inappropriate access as well as the laws and regulations that may govern the use and retention period of their data.

What are the responsibilities of the data owner?

The data owner is responsible for setting up a policy to allow specific individuals to see and update the data. Usually, a person's role determines access. For example, anyone in the accounting department can view the accounting data, but only lead accounting analysts can add new accounts.

The data owner is also responsible for determining who has access to the data, how the data should be secured, how long the data should be retained, what the appropriate disposal methods are, and whether the data should be encrypted.

The data owner may appoint an administrator to do the daily tasks associated with these responsibilities. For example, the data owner may appoint someone to approve daily requests to access the data. The appointed person will act under the direct instructions of the data owner.

Unfortunately, IT often ends up being the de facto owner of the data. Although the IT department can be the custodian of the data, it should not be the owner. Employees in IT generally do not know how important the data is to the business, how the data is to be used, and which people (or roles) should access the data.

Another reason IT should not be the owner as well as custodian of the data is separation of duties. If IT decides who has access to the data and then administers that access, there are no checks and balances to ensure that access control policies are being followed or that inappropriate access is not being assigned.

IT's role is usually that of data custodian. The custodian is responsible for implementing the policies set by the data owner. For example, IT is usually responsible for ensuring that the database files access controls (such as *PUBLIC authority) are set per the data owner's requirements. IT is also responsible for backing up the data as well as properly disposing of any electronic copies of the data in the department's possession.

How do I get the owner to take ownership?

Getting the appropriate person to take ownership of the data can be difficult, if not impossible, without upper management's involvement in the initiative. When the organization understands the importance of appointing data owners, it will understand that data is a vital business asset that must be handled properly. Having IT appoint data owners without upper management buy-in is rarely successful. So, from an IT perspective, the first order of business is to educate management on the idea that the organization's data is a vital asset and needs correct classification as well as an appropriate (non-IT) owner so that it can be handled and protected properly.

Summary

Data classification and appropriate data ownership are key elements in an organization's security policy. Without these elements, implementing a security scheme will be difficult, and an organization is unlikely to meet the internal and regulatory requirements related to access control for its data.

Carol Woodbury *is President and co-founder of SkyView Partners, Inc., a firm specializing in security policy and compliance management software as well as security consulting and remediation services. Carol has over 16 years in the security industry, 10 of those working for IBM's Enterprise Server Group as the AS/400 Security Architect and Chief Engineering Manager of Security Technology. Carol's second book, Experts' Guide to OS/400 and i5/OS Security, is available at www.amazon.com .*

Carol can be reached at carol.woodbury@skyviewpartners.com