Realtime publishers

Data Protection and Compliance in Complex Environments

The CSO Executive Series

# Finding, Classifying and Assessing Data in the Enterprise

sponsored by

CREDANT®

We Protect What Matters

Kevin Beaver

# Introduction to Realtime Publishers

**by Don Jones, Series Editor**

For several years now, Realtime has produced dozens and dozens of high-quality books that just happen to be delivered in electronic format—at no cost to you, the reader. We've made this unique publishing model work through the generous support and cooperation of our sponsors, who agree to bear each book's production expenses for the benefit of our readers.

Although we've always offered our publications to you for free, don't think for a moment that quality is anything less than our top priority. My job is to make sure that our books are as good as—and in most cases better than—any printed book that would cost you $40 or more. Our electronic publishing model offers several advantages over printed books: You receive chapters literally as fast as our authors produce them (hence the "realtime" aspect of our model), and we can update chapters to reflect the latest changes in technology.

I want to point out that our books are by no means paid advertisements or white papers. We're an independent publishing company, and an important aspect of my job is to make sure that our authors are free to voice their expertise and opinions without reservation or restriction. We maintain complete editorial control of our publications, and I'm proud that we've produced so many quality books over the past years.

I want to extend an invitation to visit us at http://nexus.realtimepublishers.com, especially if you've received this publication from a friend or colleague. We have a wide variety of additional books on a range of topics, and you're sure to find something that's of interest to you—and it won't cost you a thing. We hope you'll continue to come to Realtime for your educational needs far into the future.

Until then, enjoy.

Don Jones

## Copyright Statement

© 2010 Realtime Publishers. All rights reserved. This site contains materials that have been created, developed, or commissioned by, and published with the permission of, Realtime Publishers (the "Materials") and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of Realtime Publishers its web site sponsors. In no event shall Realtime Publishers or its web site sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

The Materials (including but not limited to the text, images, audio, and/or video) may not be copied, reproduced, republished, uploaded, posted, transmitted, or distributed in any way, in whole or in part, except that one copy may be downloaded for your personal, non-commercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Realtime Publishers and the Realtime Publishers logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners.

If you have any questions about these terms, or if you would like information about licensing materials from Realtime Publishers, please contact us via e-mail at info@realtimepublishers.com.

# Finding, Classifying, and Assessing Data in the Enterprise

It's one thing to have a set of policies and a formal security controls framework in place, but quite another to actually make things happen. You need to know not only what data you've got in order to keep it protected but also where to look for it, how to classify it, and how to determine if it's actually at risk.

## Finding Data

Your first task in gaining control of your environment is to seek out the data that needs to be protected. This task will require an inventory of critical business systems as well as knowledge of how data flows through each system. Trying to determine what's where all by yourself will likely prove to be an exercise in futility. Instead, you're going to need to collaborate with the right business managers and data administrators in order to get what you need. A simple approach to getting things kicked off is the following:

1. Ask your IT staff which servers, applications, and databases they consider critical to the business. Defining what "critical" means in your business environment will help as everyone seems to have their own definition.
2. Determine who the managers and/or administrators are for these critical systems.
3. Meet with—or send questionnaires to—those in charge and ask:
    a. What data is created, stored, or otherwise processed by each system?
    b. Can you elaborate on the data life cycle from creation to destruction?

**Are You Sure You've Thought of Everything?**

Never forget the information security principle that complexity breeds risk. In today's complex environments, it's easy to overlook something. Even when data managers and administrators tell you about everything they've got, ask more questions as they'll likely remember additional areas where data is held.

    Some data managers or administrators may already have such an inventory or documentation. If so, have them review it because odds are things have changed in the data inventory or flow since the last revision.
4. Pull together your findings and then review the inventory and prioritize the data with your security committee and/or other key stakeholders such as audit, compliance, and legal.

No doubt this is a simplified view of the data inventory process, but a proven one nonetheless. Also, keep in mind that these steps could've already been performed—albeit out of date—during a previous business impact analysis or data backup inventory. Even if it's something that was done years ago before you came on board or your position was created, it could still be of great benefit.

**Don't Overlook the *Not-So-Obvious***

The last thing you want to have happen is a security breach and then one of your data owners saying "Oh yeah, I forgot about that data." Sad but true—it happens all the time. Be sure to consider non-traditional locations for data such as:

- Desktop computers (It's important to note that many people assume that because they store everything on the network that there's nothing of value on their laptops—not true. There's email, passwords, Web browser caches, and files such as word processing documents, PDFs, and spreadsheets that are undoubtedly stored locally.)
- Laptop computers (Like desktop computers, a lot of valuable data resides on laptops, and as the workforce becomes more mobile, such is becoming ever more the case. Consider, for example, a nurse making house calls; she brings with her a laptop that has protected health information for multiple patients.)
- Mobile devices including tablet computers, netbooks, smartphones, and iPods or other external hard drives, many of which may be synchronizing data from the network or the employee's work computer
- External media such as USB thumb drives (Many of your best employees bring valuable data out of the office on these devices, possibly to work on during evenings and weekends.)
- Data backups including tapes, disk images, CDs, DVDs, and removable media
- Data in transit that may merely flow through some of your systems and not get stored on the network
- Wi-Fi related systems including guest networks
- Databases including those associated with intranet sites
- Development and test systems where production data are often present

## Classifying and Assessing Data

Now that you know what's where, the next step is to classify your data and see how it's at risk. Before you get started, it's important to remember the KISS (Keep It Simple Stupid) method of classification. If you create an intricate matrix of various data types, the likely outcome is increased information systems complexity. Data management and security are difficult enough, so avoid this common practice wherever possible. Keeping in mind that every situation is different, a practical approach is to separate data into the following categories:

- Public—Data that can be disclosed openly to the general public such as:
    - General Web site marketing data
    - Business partner lists
    - Regulatory statements of compliance
- Internal Use Only—Data that should only be made accessible to employees or other authorized parties but wouldn't necessarily harm an individual or the business if it were disclosed such as:
    - Email distribution lists
    - Staff evaluations
    - Internal security policies
- Confidential—Data that is highly sensitive and would lead to the most detrimental loss if it were disclosed, compromised, or exploited, such as:
    - Intellectual property such as source code and engineering diagrams
    - Employee Social Security numbers
    - Customer financial records

Unless you work in an extremely large and complex business environment—or for the military—these three data classifications are likely to serve your business well for any rational data protection and compliance needs.

**Not All Data Are Created Equal**

All too often the assumption is that every system, application, and data set needs the same level of protection across the board. This approach is not only costly but also sets you up for failure by trying to be all things to all people. Rather than taking a one-size-fits-all approach to data protection, it would behoove you to focus on the critical data that matters most to the business. If you take the proven time-management approach of focusing on systems and data that are both urgent and important, you'll be able to fine tune your protective measures, save money, and not drive yourself crazy all at the same time.

Another consideration for data classification is business value versus compliance and regulatory value. Is one data set more important than the other? Is there more to lose if, say, healthcare records covered under HIPAA or the HITECH Act were compromised as opposed to intellectual property? These are difficult questions that must be answered. This example underscores the importance of having a good team of people to work with such as audit, compliance, and legal to ensure all the right things are being considered and the best possible decisions are being made.

Once you've classified different data types, the final step is to determine how exactly it's being put at risk. The risks you uncover will guide your overall strategy and ultimately determine specific controls necessary for ongoing data protection and compliance.

**It's All About Risk**

The overall concept of information risk is something you need to use in every IT and security-related decision you make. Risks are either unacceptable and move the business away from its goals or acceptable and help move the business toward its goals. It's up to you and your team of stakeholders to decide which ones count and which ones don't.

Certain people believe that quantitative risk analysis (a practice where you mathematically calculate the rates of occurrence and loss expectancies) is most beneficial. However, in keeping with the KISS method, I strongly believe you can perform a *qualitative* analysis (a practice where you generally rank the likelihood of a risk occurring and the overall impact it would have on the business) and still effectively manage information risks. Table 1 shows a simplified representation of a qualitative risk analysis.

| | High Likelihood | Medium Likelihood | Low Likelihood |
|---|---|---|---|
| **High Impact** | Unencrypted customer data stored on external hard drives used at sales conferences | Tablet computers storing R&D documents that aren't being properly backed up | Weak password on an internal server that houses source code |
| **Medium Impact** | Corporate emails sent over unsecured wireless networks at hotels and meeting facilities | Lack of endpoint controls on IT computers that house security policy documents | Exploitable missing patch on an internal database system housing employee evaluations |
| **Low Impact** | Lack of current antivirus software on standalone training computers | Security guards browsing the Internet using their personal laptops on the guest network | Poorly-configured software exposing data in transit to/from the marketing Web site |

**Table 1: A qualitative analysis using likelihood and impact to determine risk.**

## Tools and Technologies Available

Security tools can be your friend, especially when it comes to finding what data you have and, in turn, keeping it under wraps. There are specific technologies that can help you bring things full circle such as:

- Seeking out sensitive data on the network, which can complement manual searches and inventories or even eliminate the need for manual searches altogether
- Classifying the data based on your own classification rules
- Determining how data is at risk by pointing out context-based vulnerabilities

Good tools not only serve to protect data but will also help with your ongoing compliance needs. They lend themselves to scenarios whereby you can proactively hand over audit and compliance reports rather than wait for someone else to tell you where things stand. Perhaps most importantly, properly-implemented tools allow you to reach a level of automation, visibility, and control you'd be unable to attain otherwise.