CorporateGovernor series:
Covering your assets: A proactive approach to securing sensitive data

Addressing the concerns of the Corporate Governance Community Winter 2012
## Contents

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Tighter regulations</td>
</tr>
<tr>
<td>3</td>
<td>Scoping the data landscape: Where is your sensitive information?</td>
</tr>
<tr>
<td>4</td>
<td>Important steps in the data discovery process</td>
</tr>
<tr>
<td>7</td>
<td>Data leakage: A case study</td>
</tr>
<tr>
<td>8</td>
<td>About the authors</td>
</tr>
<tr>
<td>9</td>
<td>Grant Thornton offices</td>
</tr>
</tbody>
</table>
Introduction

High-profile cyberattacks and breaches of information assets are common news items lately. As some companies suffer through public relations consequences and loss of business while restoring services, the issue of data security has risen in importance for all organizations. While protecting trade secrets and confidential information has always been considered a vital business activity, industry standards and state and federal regulations concerning data protection and privacy are growing increasingly stringent. Now more than ever before, organizations understand that failure to comply has real consequences, including fines and litigation. Without a plan for effective security, companies can experience productivity losses, customer defections and reputational damage.
In recent years, the federal government and some states have responded to heightened concerns about protecting various types of sensitive data, including personally identifiable information (PII), protected health information (PHI) and credit card data. New regulations and standards put the burden on the organization to comply with rules and to shield the sensitive data it uses from unauthorized disclosure. In addition to the United States, the European Union and some Asian countries have created very specific data protection regulations.

Following are examples:

• The Health Insurance Portability and Accountability Act (HIPAA) originally became federal law in 1996. Partly in response to the rising incidence of medical identity fraud, the government passed the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009 as part of that year’s stimulus plan. The HITECH Act focuses on protecting sensitive health information from unauthorized access; such protection is particularly important given the growing use of electronic medical records.

• Individual U.S. states have responded to data breaches and consumer concerns by passing legislation governing the protection of PII such as Social Security numbers and driver’s license numbers.

• Five major credit card networks have formed the Payment Card Industry Security Standards Council (PCI SSC) to establish industry standards for the protection of cardholder data. The primary standard produced by the PCI SSC — the Payment Card Industry Data Security Standard (PCI DSS) — has emerged as the broadest and most prescriptive compliance initiative related to the safeguarding of personal data.

The Sarbanes-Oxley Act addresses the integrity of company financial data. In contrast, HIPAA, HITECH, state privacy laws and the PCI DSS are focused on the confidentiality of sensitive consumer data. These standards are data-centric in that they are designed to protect sensitive data associated with individuals; moreover, their application depends largely upon the location and flow of the data.
The first step in data security compliance is to establish the scope of the task: Where does the data enter the organization? Where does it flow? And finally, where does it reside? This sounds elementary, but despite well-publicized incidents of data theft, companies often lack specific knowledge of where sensitive data resides and how it flows through their IT infrastructure. In addition, the more locations in which sensitive information is stored, the more challenging its protection becomes.

There is the risk that sensitive data may reside in unknown locations. This suggests that the very existence of the data may be unknown to those responsible for its security. Indeed, the discovery phase encompasses not only the locations that the organization knows about, but also informal, undocumented locations. For example, credit card data may be found in screenshots that are sent to the help desk or in spreadsheets that are used to analyze consumer purchasing patterns. Items like these that contain sensitive information (e.g., Social Security numbers, credit card numbers) may be discovered to have been stored indefinitely in a departmental network folder without sufficiently strict permissions — and network folders, even those containing sensitive information, are typically included in daily, weekly and monthly backup tapes. To help prevent the possibility of a security breach, an organization needs to develop comprehensive data scoping processes so that it can determine all locations where sensitive data is stored.
Throughout the data discovery phase, information is gathered in the context of the compliance effort — that is, within the framework of requirements such as HIPAA and PCI DSS that govern PII. There are four important steps in the discovery process:

1. Identifying the entry points for sensitive data
2. Capturing the life cycle of sensitive data by documenting the detailed data flows through the time that the data exits the system
3. Mapping how the data flows through the IT infrastructure of the enterprise
4. Validating findings with automated tools such as data leakage prevention (DLP) applications

Whatever is driving the discovery process, the goal is to answer these questions:
- Where does the data originate?
- Where is the data stored?
- How is the data used?
- Who has access to the data?
- How and at what point is the data destroyed?

Organizations should take the following steps to understand the location and flow of sensitive data.

**Step 1: Identify the entry points for sensitive data**
The initial objective of the discovery process is to identify all the points where sensitive data can enter the organization — in other words, those points at which the organization accepts or collects the information. This knowledge is gleaned through interviews and reviews of documentation.

From a PCI DSS perspective, an entry point for a credit card merchant is the point of purchase — when the customer swipes his or her credit card number or enters his or her credit card information on the merchant’s e-commerce website. From a HIPAA/HITECH perspective, an entry point may be in a hospital setting when admissions personnel collect medical history and insurance information, or when a doctor enters treatment details into patient records. From a privacy perspective, an entry point may be the point at which an insurance company or financial institution collects Social Security numbers from clients.

**Step 2: Capture the life cycle of sensitive data by documenting the detailed data flows through the time that the data exits the system**
The data life cycle is an essential principle of data security and datacentric compliance. HIPAA and other regulations set forth requirements for data retention within a specified time period. But retaining sensitive data indefinitely is neither practical nor low risk.

**Step 3: Map how the data flows through the IT infrastructure of the enterprise**
The organization needs to document the data flows within its IT infrastructure. By determining how data is flowing through the enterprise, the organization will be able to identify the computers, servers and networks that are critical to compliance — in other words, those that enable the collection, flow and storage of sensitive data.
components that store, process or transmit sensitive data. In this way, it is possible to use discretion in identifying which segments of the IT architecture warrant monitoring. It is also possible to leverage DLP to discover additional information flows involving those segments — and to find previously undocumented storage locations. Taking this approach avoids the need for the organization to undergo a more costly, time-intensive process that may yield questionable results.

In order for the DLP tool to perform effective discovery, the organization must identify the criteria that define sensitive data. For certain data types, these criteria are relatively well defined and well documented. Credit card numbers, for instance, should follow a 16-digit format and pass the Mod 10 check, which is sometimes referred to as the Luhn formula, after its inventor. Other types of sensitive data may be more difficult to define if the discovery process is to avoid having many false-positives. For example, when querying for Social Security numbers, the user might define the search criteria based on a nine-digit number. However, this search will likely identify many extraneous items, and unfortunately, Social Security numbers don’t have an equivalent to the Luhn formula that would reduce the likelihood of false-positives.

**Step 4: Validate findings with automated tools such as DLP applications**

As their name implies, DLP tools are designed to prevent the unwanted leakage of certain sensitive data from the organization. Depending on how they are used, DLP tools can identify incidents of unintentional data use, transmission or storage. There are two common modes of use for DLP tools: discovery and prevention.

Once the organization has performed its initial discovery and mapping process, DLP tools in discovery mode can be used to validate these findings. DLP technology will identify where the data is stored, how it flows through the organization’s IT infrastructure, and where it interacts with users. When used in discovery mode, a DLP tool identifies only the location and flow of the sensitive data that it discovers. In prevention mode, DLP tools can be programmed to stop unwanted information flow that violates a specified policy.

Once an organization has developed a baseline understanding of the locations and flow of its sensitive data, DLP tools can be used as a broad preventive control to head off unauthorized traffic or disclosure of that data.

The sidebar at the right lists a number of commercial and open-source DLP tools that are available. The open-source tools typically do not scale or perform as well as the commercial offerings do, but they can operate effectively if the organization is highly focused on certain data sets that require scrutiny.

**Using DLP technology**

It is neither practical nor wise for an organization to use DLP technology for discovery purposes at every storage location or server or on every network component. Leveraging knowledge gained through the manual discovery and mapping process, the organization should focus its efforts on the identified IT components that store, process or transmit sensitive data. In this way, it is possible to use discretion in identifying which segments of the IT architecture warrant monitoring. It is also possible to leverage DLP to discover additional information flows involving those segments — and to find previously undocumented storage locations. Taking this approach avoids the need for the organization to undergo a more costly, time-intensive process that may yield questionable results.

In order for the DLP tool to perform effective discovery, the organization must identify the criteria that define sensitive data. For certain data types, these criteria are relatively well defined and well documented. Credit card numbers, for instance, should follow a 16-digit format and pass the Mod 10 check, which is sometimes referred to as the Luhn formula, after its inventor. Other types of sensitive data may be more difficult to define if the discovery process is to avoid having many false-positives. For example, when querying for Social Security numbers, the user might define the search criteria based on a nine-digit number. However, this search will likely identify many extraneous items, and unfortunately, Social Security numbers don’t have an equivalent to the Luhn formula that would reduce the likelihood of false-positives.

**Data discovery tools**

- Commercial
  - CA DLP
  - Symantec Data Loss Prevention
  - Fidelis XPS™
  - McAfee Data Loss Prevention
  - RSA Data Loss Prevention
  - Trend Micro® LeakProof™
  - Verdasys Digital Guardian
- Open source
  - Websense Data Security Suite
  - Spider
  - Cscore
  - Senf
  - FTimes
  - Snort®
Whether DLP tools are being used for discovery or prevention, they encounter sensitive data in one of three states:

- **In motion** as it flows through the organization’s IT infrastructure via network protocols such as Hypertext Transfer Protocol (HTTP), File Transfer Protocol (FTP) and Simple Mail Transfer Protocol (SMTP). Typically, the DLP tool works in concert with a network sensor or appliance that is placed in a position to see network traffic.

- **At rest** when it is stored in servers or on a storage area network. In this situation, the DLP tool typically takes the form of a software agent that runs on the platform, much like anti-virus software.

- **In use** when it is located within end-user hosts such as workstations or laptops. Under these circumstances, the DLP tool is installed as a software agent.

After performing an automated discovery process, the organization must eliminate false-positive. This is inherently a manual process; it cannot be automated. Working from the list of locations discovered by the DLP tool, knowledgeable personnel need to examine each data location (e.g., files, database tables) and inspect its contents to determine whether it actually contains sensitive data.

If the location does not contain sensitive data, it should be marked as a false-positive. If it contains authorized sensitive data — in other words, if the organization has been aware that sensitive data is stored there — then the data should be marked as valid. If the location contains unauthorized sensitive data — that is, if the organization was previously unaware that sensitive data was stored there — then the organization should investigate further to determine whether storage of the data is permissible in that location.

**What’s next?**
More often than not, discovery reveals a broad swathe of sensitive data that is accessed, used and stored on servers, networks and user platforms across the enterprise. Given the extensive reach of this data, compliance with HIPAA or the PCI DSS can become a very costly and resource-intensive activity. However, there are several common-sense approaches that companies can adopt to reduce the reach of sensitive data and the potential for unauthorized traffic or disclosure.
**Data leakage: A case study**

**Background**
A financial institution with $3.7 billion in assets and full-service branches throughout the United States was increasing its reliance on storing information electronically. The institution wanted to know exactly where confidential client data was stored and whether that data was easily accessible by intruders. Storing unencrypted credit card data is a PCI DSS violation, and storing any unencrypted PII presents a scenario where personal information might be misused.

**Problem**
Dissatisfied with its current security service provider after a data breach, the institution sought a fresh perspective on its network and application security risk profile. Management requested assistance in determining whether confidential client data was being stored on the institution’s servers without being encrypted.

**Solution**
Over the next 30 days, a Grant Thornton LLP team conducted an inventory of confidential client data. A network content scan using a DLP tool identified files and database tables that potentially contained credit card numbers or other PII. The Grant Thornton team worked with the institution’s IT staff to:

- locate where confidential client data was stored;
- identify the entry points for this data;
- determine whether confidential data was unencrypted, how it was used, and who had access to it; and
- document when and how confidential data was archived and destroyed.

**Results**
A network scan of the institution’s 25 servers revealed that 538 files potentially contained unencrypted credit card data and 3,599 files potentially contained unencrypted PII. This came as a surprise to the chief information officer (CIO) because documented business processes precluded the storage of credit card data or other PII without encryption on the institution’s servers. With our help, the institution took the following actions over a 60-day period:

- Removed credit card data from the identified files
- Instructed the branches not to store any credit card data or other PII unless this information is required to complete the business function
- Evaluated DLP tools that quarantine or flag sensitive data
- Implemented processes to ensure that diagnostic files are immediately removed from the server after troubleshooting exercises have taken place
- Confirmed compliance with appropriate regulatory requirements

**Recommendation**
Failure to protect sensitive information can lead to security breaches and unauthorized disclosure of customer data. This can trigger regulatory sanctions, financial penalties, notification and monitoring costs, reputational damage and a shrinking customer base. Therefore, organizations need to adopt a strategic but practical risk-based approach to the use and protection of sensitive data. This has broader benefits than simply the maintenance of data security; it can also support better business decision-making by limiting the storage of redundant data and eliminating the reconciliations and rework that often result from such redundancy.

---

1 Confidential client data includes account numbers — such as credit card numbers — and other PII. PII is information that can be used to uniquely identify, contact or locate a person.
Brian Browne is a senior manager in the Business Advisory Services practice within the Philadelphia office of Grant Thornton LLP. He is also the practice leader of the firm’s Cybersecurity group in the Northeast Region. Browne has more than 20 years of experience with clients in all industries and is skilled in all aspects of information security, from management through hands-on implementation. His information security experience encompasses the requirements of the Payment Card Industry Data Security Standard, the Health Insurance Portability and Accountability Act, and the Sarbanes-Oxley Act.

Danny Miller is a principal in the Business Advisory Services practice within the Philadelphia office of Grant Thornton LLP. He is the national practice leader of the firm’s Cybersecurity group. Miller has comprehensive experience in a variety of industries, including higher education, not-for-profit, financial services, banking, oil, gas, consumer and industrial products, construction, real estate and manufacturing. He has broad experience in those same industries internationally, including in European Union countries and Southeast Asia. Miller has more than 25 years of experience in auditing, business advisory services and information security.
About Grant Thornton’s Advisory Services Practice

Today you need advisors who focus on insightful and innovative solutions for your complex issues, such as complying with changing legislation, managing risk, containing costs, streamlining business processes and identifying strategic transaction opportunities. Grant Thornton’s Advisory Services professionals can deliver value by providing independent advice to public, private and not-for-profit organizations. Our specialists combine insight and innovation from multiple disciplines with a wide range of business and industry knowledge. To learn more, visit www.GrantThornton.com/advisory.