Analyst Insight

Aberdeen Group
A Harte-Hanks Company

February 2013

Three Steps to Successful Data Classification

Segmenting — or **classifying** — your organization's confidential information and intellectual property is a cornerstone for protecting it. Data classification helps to ensure not only that the appropriate levels of policies, controls, and resources are in place, but also that these investments deliver an appropriate level of value to the business in return. Three core steps to successful data classification initiatives include **creating a data classification scheme**; **empowering end-users** through ongoing awareness and training; and **communicating the results and trends** of a regular review and analysis to the owners of the business risks for the confidential information and intellectual property that is being protected.

Analyst Insight

Aberdeen's Analyst Insights provide the analyst perspective of the research as drawn from an aggregated view of surveys, interviews, analysis, and industry experience.

Business Context: Management's Discussion of Risk

For publicly traded companies in the United States, some insights about the degree of focus and the level of sophistication their executive leadership gives to managing their most critical enterprise risks are available through their standard Securities and Exchange Commission (SEC) Form 10-K filings, which feature management's discussion of risk factors that could materially affect the company's business, operations, or financial condition.

Nearly four years ago, Aberdeen observed that IT-related risks were not even showing up in the SEC 10-K filings for leading US high-tech firms (see *IT GRC: Managing Risk, Improving Visibility, and Reducing Operating Costs*, May 2009). As noted then:

- "As an illustration, the risk factors from the most recent 10-K filing for a US-based high-tech company with greater than $10 billion in annual revenue are summarized … [and grouped] into four high-level categories: *financial*, *strategic*, *operational*, and *other*. Look closely … and notice that not once is Information Technology (IT) mentioned directly. Yet IT plays a fundamental, foundational role in addressing many of these key risk factors, and indeed IT is responsible for supporting or enabling numerous aspects of any given company's business."

"Security breaches and other disruptions to the Company's information technology infrastructure could interfere with the Company's operations, compromise information belonging to the Company and its customers and suppliers, and expose the Company to liability which could adversely impact the Company's business and reputation."

~ 3M, SEC Form 10-K, 2012

Has this situation changed over time? Four years later, Aberdeen's review and analysis of Section 1A (Risk Factors) from the most recent SEC Form 10-K filings from the 30 publicly traded companies that make up the Dow Jones Industrial Average reveals some interesting patterns:

- Twenty-five out of 30 companies commented on the risks and consequences of the compromise of **confidential information** — information belonging not only to the company, but also to its business partners and customers (see Figure 1, column "CI")

"The Company is increasingly dependent on sophisticated information technology and infrastructure. Any significant breakdown, intrusion, interruption or corruption of these systems or data breaches could have a material adverse effect on our business."

~ Merck, SEC Form 10-K, 2012

- But just nine out of 30 companies explicitly discussed the risks related to the security of their **intellectual property** and trade secrets (see Figure 1, column "IP")

Figure 1: Analysis of 2012 SEC Form 10-K Filings of the Dow Jones Industrial Average (30 firms)

Symbol	Company	CI[1]	IP[2]	IT[3]	3P[4]	Frequency Analysis of Section 1A (Risk Factors)
DIS	Walt Disney	X	X	X	X	
JNJ	Johnson & Johnson	X	X	X		
MSFT	Microsoft	X	X	X		
UTX	United Technologies	X	X	X		
BAC	Bank of America	X		X	X	
GE	General Electric	X		X	X	
PFE	Pfizer	X		X	X	
PG	Procter & Gamble	X		X	X	
KO	Coca-Cola	X		X	X	
TRV	Travelers Companies	X		X	X	
UNH	UnitedHealth	X		X	X	
MMM	3M	X		X		
T	AT&T	X		X		
XOM	Exxon Mobil	X		X		
HPQ	Hewlett-Packard	X		X		
JPM	JPMorgan Chase	X		X		
MRK	Merck	X		X		
VZ	Verizon	X		X		
WMT	Wal-Mart	X		X		
INTC	Intel	X	X			
MCD	McDonald's	X	X			
BA	Boeing	X				
CAT	Caterpillar	X				
HD	Home Depot	X				
IBM	IBM	X				
AA	Alcoa		X			
AXP	American Express		X			
DD	Du Pont		X			
CVX	Chevron			X		
CSCO	Cisco Systems			X		

Note 1: Compromise of confidential information; *Note 2*: Compromise of intellectual property
Note 3: Disruption of operations from reliance on IT; *Note 4*: Disruption of operations from reliance on third party IT
Data: 2012 SEC Form10-K Filings of the Dow Jones Industrial Average companies
Source: Aberdeen Group, February 2013

A word cloud created from a frequency analysis of the 30 10-K discussions for security-related risks (Figure 1) provides some high-level insights into the current *language* of risk from the executive management point of view. Most importantly, it provides evidence of a positive trend: that senior management's understanding of these risks is slowly but surely going mainstream. (The SEC Division of Corporation Finance guidance regarding disclosure obligations relating to cyber security risks and incidents, published in October 2011, was undoubtedly a major catalyst behind this change in the 10-K filings for 2012.)

A more detailed review and analysis of the SEC 10-K discussions shows that leading companies are starting to bridge an important language gap in their discussions of risk by making a business connection between the *unrewarded* risks of security and compliance and the *rewarded* risks of operations, innovation, and growth (see Table 1).

Table 1: Bridging the Language Gap Between Two Types of Risk

	Unrewarded Risk	Rewarded Risk
Risk Management Objectives	▪ Protect value ▪ Defend assets ▪ Minimize downside	▪ Create value ▪ Enable assets ▪ Maximize upside
Example Areas of Focus	▪ Security vulnerabilities and threats ▪ Regulatory compliance	▪ Innovation and growth initiatives ▪ Operational efficiencies
Associated Assets	▪ Identities and access ▪ Applications, data, IP ▪ IT infrastructure ▪ Physical infrastructure ▪ Personnel safety	▪ Revenue streams ▪ Distribution channels ▪ Products and services ▪ Operations and supply chain ▪ Reputation and brand

Source: Aberdeen Group, February 2013

Progress has clearly been made, but important gaps in the discussions about risks —to **intellectual property** and trade secrets, and from an increasing reliance on **third-party IT infrastructure** — provide evidence that security still has far to go as a board-level issue. This point is amplified when one recalls that the companies in the Dow Jones Industrial Average represent the generally more mature, sophisticated tip of the overall market pyramid.

The Role of Data Classification in Protecting Confidential Information and Intellectual Property

Over the last six years, Aberdeen's research has consistently shown that the following general steps are among those correlated with top performance at safeguarding confidential information and intellectual property:

Fast Facts

√ Twenty-one out of 30 companies commented on the risks and consequences of disruptions in their IT infrastructure (see Figure 1, column "IT")

√ But just 8 out of 30 companies explicitly discussed the risks from their increasing reliance on third-party infrastructure, e.g., managed services and cloud service providers (see Figure 1, column "3P")

"Inability to protect and enforce the company's intellectual property rights could adversely affect the company's financial results. Intellectual property rights, including patents, plant variety protection, trade secrets, confidential information, trademarks, tradenames and other forms of trade dress, are important to the company's business.

The company has designed and implemented internal controls to restrict access to and distribution of its intellectual property. Despite these precautions, the company's intellectual property is vulnerable to unauthorized access through cyber-attacks, theft, and other security breaches."

~ E. I. du Pont de Nemours and Co., SEC Form 10-K, 2012

- **Identify and classify your data** — you can't manage data that you don't know about, and not all data is worth the same level of protection
- **Prioritize your security control objectives** for these information assets as a function of risk, audit, and compliance requirements (another way to think about data classification)
- **Establish consistent policies** as part of an overall approach to safeguarding sensitive data, wherever it may flow — at rest in the back-end, in motion on the network, and in use at the endpoints

As Aberdeen has noted many times, there is a growing appreciation that "back-end systems" no longer refers only to the networks, hosts, storage, and applications within the enterprise server rooms or datacenters — it also refers to virtualized IT infrastructure, whether on premise or in the "*cloud*." Similarly, "networks" no longer refers only to the electronic interconnections and communications protocols between systems — it also refers to *social* connections and collaboration between people. And "endpoints" no longer refers only to the devices that are centrally procured, provisioned, and managed by the enterprise IT function — it also refers to the distributed and highly *mobile* devices that are increasingly procured, provisioned, and managed by enterprise end-users themselves.

What this means is that the confidential information and intellectual property your organization is trying to protect is increasingly likely to be flowing from back-end systems that it doesn't control, through networks that it doesn't control, to endpoints and end-users that it doesn't control.

Step 1: Creating a Data Classification Scheme

Segmenting — or **classifying** — your organization's confidential information and intellectual property is a cornerstone for protecting it. Classification helps to ensure not only that the appropriate levels of policies, controls, and resources are in place, but also that these investments deliver an appropriate level of value to the business in return. Classification helps companies protect the data that is worth protecting, and do so in a cost-effective, risk-based manner.

Even organizations that have not yet implemented data classification schemes are most likely familiar with what they look like — for example, the classic **unclassified**, **secret,** and **top secret** classifications used in government and defense, or designations such as **public**, **proprietary** or **internal use only,** and **company confidential** found in commercial enterprise settings (see Table 2). Some general guidelines for data classification schemes include the following:

- Classifications should apply regardless of the format of the information (e.g., electronic, paper, recordings, applications)
- Classifications should be unique and distinct (no overlaps)

Definitions

√ In its simplest form, **data classification** distinguishes between data which requires protection and that which does not

√ **Levels** of classification correspond to degrees of protection and controls

Data Classification Schemes

Basic elements include:

√ Definition of classification levels

√ Criteria to determine how information is to be classified

√ Controls required for each classification, including level of assurance required for end-user authentication; rules or roles governing end-user access; protections to implement appropriate levels of confidentiality, integrity, and availability

√ Documentation for exceptions to any of the above

√ Definition of responsibilities, e.g., data owners, data custodians, and processes for transferring responsibilities from one owner or custodian to another

√ Periodic audits of classification and ownership, and processes for remediation in the case of errors or inconsistencies

√ Ongoing documentation, awareness, and training for all information users, to make them aware of their responsibilities for handling data at each level of classification

- Classifications should be neither too many (which is likely to be confusing) nor too few (which may give the impression of being of too little importance or consequence)
- Classifications should be clear and unambiguous (e.g., some organizations do not favor designations such as "PUBLIC" because it might inadvertently imply that such information is appropriate for sharing with press, analysts, and in social media)
- Classifications should be kept at the top level, with the use of appropriate *caveats* for other important concepts, for example:
    - PUBLIC
        - Apply Markings: <NO, YES>
    - INTERNAL USE ONLY
    - CONFIDENTIAL
        - Sensitivity: <CONTAINS TRADE SECRETS, etc.>
        - Project: <list of projects>
    - RESTRICTED
        - Project: <list of projects>
- Definition of acceptable *Distribution List* or *Releasable To* caveats can also be useful (e.g., to designate "PROPRIETARY" information that is allowed to be shared with select customers or business partners)

Table 2: Data Classification Levels for Commercial, Government and Defense (illustrative)

Level	Commercial	Government	NATO[1]
Lowest	▪ NOT SENSITIVE ▪ NOT CLASSIFIED ▪ NON-PROPRIETARY ▪ PUBLIC	▪ UNCLASSIFIED	▪ NATO UNCLASSIFIED
	▪ PROPRIETARY ▪ INTERNAL USE ONLY	▪ CONFIDENTIAL	▪ NATO RESTRICTED
	▪ CONFIDENTIAL	▪ SECRET	▪ NATO CLASSIFIED
Highest	▪ RESTRICTED ▪ HIGHLY CONFIDENTIAL	▪ TOP SECRET	

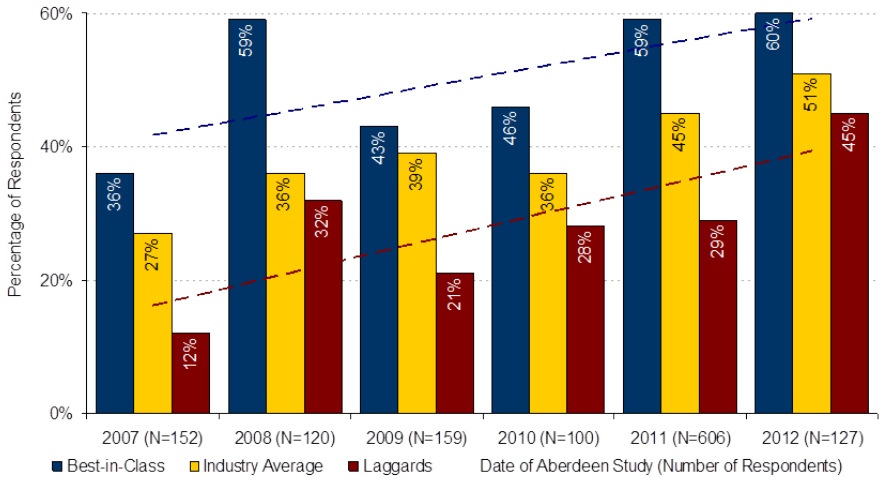
Note 1: e.g., NATO INFOSEC Directive AC/322
Source: Aberdeen Group, February 2013

Best Practices: Aberdeen's Research Findings

In six separate studies on data protection conducted over the last six years, Aberdeen's research has shown that **data classification is consistently correlated with the achievement of top performance** (Figure 2):

- Compared to the lagging performers, the leading performers in each study are *between 1.5-times and 3-times more likely* to indicate that data classification is a current capability
- The average year-over-year growth in companies implementing data classification over this period has been between 4% and 5%
- The average for all respondents who have implemented data classification has now passed 50%, a sign of its continuing maturity

Figure 2: Data Classification as a Current Enterprise Capability, by Maturity Class, in Six Independent Studies Between 2007-2012



Source: Aberdeen Group, February 2013

Aberdeen's Maturity Classes

To distinguish **Best-in-Class** (top 20%) companies from **Industry Average** (middle 50%) and **Laggard** (bottom 30%) organizations in the area of data protection, Aberdeen generally uses performance criteria such as the estimated number of the following incidents actually experienced in the last 12 months, along with the estimated year-over-year change in *unauthorized access*, *audit deficiencies*, and *data loss or exposure*. Full details are available in each respective study (see *Related Research*).

Step 2: Empowering End-Users; Transforming Behavior

Even the most carefully crafted classification schemes are unlikely to succeed if the organization does not involve its information creators in the classification of confidential information and intellectual property as part of their standard way of doing business. Data classification technologies can be a very useful means to this end, but those that *involve* the user (e.g., an *auto-suggest* feature at the time of creation) rather than *short-circuit* the user are more likely not only to reinforce policies, but also to create a sustainable corporate culture regarding data protection over time. The seamless integration of data classification into everyday workflow is the most sustainable way to change end-user behaviors and drive continuous improvement over time.

As Aberdeen has written numerous times, **awareness and training** for end-users — whether through *standalone educational programs*, or through *real-time notification of policy violations* by way of the technical capabilities of the deployed solutions — helps not only to create a general sensitivity to security and compliance for confidential information and intellectual property, but also to foster greater accountability for the data creators to see that it is properly classified and protected.

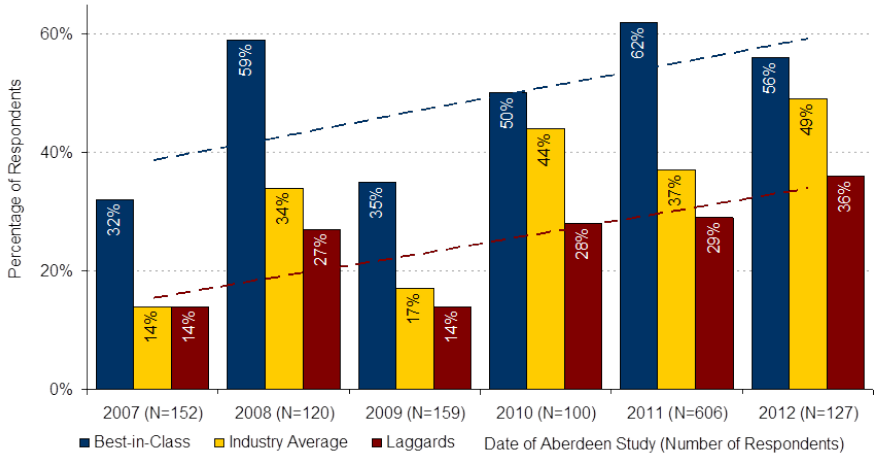
In addition, numerous Aberdeen studies on data protection support the prevailing wisdom that the majority of data loss or data exposure incidents are the result not of the *malicious* actions of outsiders, but of the *inadvertent* and *well-intentioned* actions of insiders. In other words, most incidents are the result of basic human error (*inadvertent*), and legitimate users who are simply focused on getting their jobs done (*intentional*, but not *malicious*).

Best Practices: Aberdeen's Research Findings

In the same six studies referenced above, Aberdeen's research has shown that **documentation, awareness, and training for end-users is consistently correlated with the achievement of top performance** (Figure 3):

- Compared to the lagging performers, the leading performers in each study are *between 1.5-times and 2.5-times more likely* to indicate that end-user education is a current capability
- The average year-over-year growth in companies implementing end-user education over this period has been between 5% and 6%
- The average for all respondents who have implemented end-user education is still less than 50% — part of a curious but consistent pattern that companies often spend millions of dollars on security technologies, but spend nothing on awareness and training for end-users

Figure 3: End-user Education as a Current Enterprise Capability, by Maturity Class, in Six Independent Studies Between 2007-2012



Source: Aberdeen Group, February 2013

Step 3: Closing the Loop; Reviewing and Sharing Results

As discussed previously, data classification helps to ensure not only that the appropriate levels of policies, controls, and resources are in place to protect

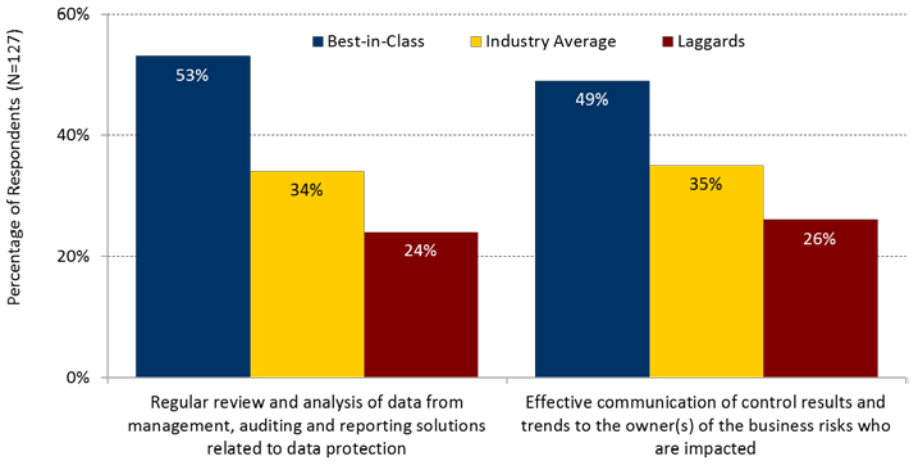
confidential information and intellectual property, but also that these investments deliver an appropriate level of business value in return.

Making this assessment is not a "one-and-done" event. It requires regular review and analysis of the information from the organization's management, auditing, and reporting systems. Just as importantly, it requires effective communication of results and trends to the owners of the business risks for the confidential information and intellectual property that is being protected. Top performers measure and monitor regularly to drive continuous improvements, e.g., by finding and eliminating root causes for incidents, excessive exceptions, or inefficient workflow.

Best Practices: Aberdeen's Research Findings

Based on the findings from the most recent of the six studies referenced above, the leading performers are more than 2-times more likely than the lagging performers to review and analyze the auditing and reporting data from their data classification and data protection solutions (Figure 4). In addition, they are nearly 2-times more likely to communicate the results and trends to the owners of the business risks, making their investments in review and analysis even more effective. If the information that's being generated isn't received and consumed — like the proverbial tree that falls in the forest, without anyone there to hear it — then it isn't making a sound.

Figure 4: Regular Review and Analysis, and Communication of Control Results and Trends to the Owners of the Business Risks



Source: Aberdeen Group, February 2013

Solutions Landscape (illustrative)

Solution providers related to identifying, classifying, and protecting confidential information and intellectual property can range from smaller specialists to multi-billion dollar firms, including the following:

- McAfee
- Symantec
- Check Point
- Cisco
- Sophos
- Trustwave
- RSA / EMC
- SafeNet
- Websense
- Trend Micro
- TITUS
- Clearswift
- BlueCoat
- Wave Systems

Summary and Key Takeaways

- Senior management's understanding of the risks related to confidential information and intellectual property is slowly but surely going mainstream. Aberdeen's analysis of Section 1A (Risk Factors) from the most recent SEC Form 10-K filings of the 30 publicly traded companies that make up the Dow Jones Industrial Average provides some interesting evidence for this positive trend.
- Leading companies are starting to bridge an important *language* gap in their discussions of risk by making a business connection between the *unrewarded* risks of security and compliance and the *rewarded* risks of operations, innovation, and growth.
- In spite of this progress, important gaps in management's discussions about risks — e.g., risks to *intellectual property* and *trade secrets* — illustrate that security still has far to go as a board-level issue.
- Ongoing IT infrastructure transformations exacerbate the challenge: the confidential information and intellectual property that organizations are trying to protect is increasingly likely to be flowing from back-end systems they don't control, through networks they don't control, to endpoints and end-users they don't control.
- **Data classification** is highly correlated with top performance at safeguarding confidential information and intellectual property, as Aberdeen's research has consistently shown over the last six years.
- Aberdeen's research confirms three fundamental steps as essential to successful data classification initiatives:
    - **Create a data classification scheme**, consistent with the high-level guidelines outlined on pages 4 and 5
    - **Empower end-users through** ongoing awareness and training; **transform behavior and culture** by integrating data classification into day-to-day workflow
    - **Close the loop** through regular review and analysis of the information from management, auditing, and reporting systems, and **communicate the results and trends** to the owners of the business risks for the confidential information and intellectual property that is being protected
- Success at identifying and classifying confidential information and intellectual property lays the necessary foundation for the achievement of Best-in-Class results, including fewer security-

related incidents, fewer compliance or audit deficiencies, and the ability to support higher scale at lower total cost.

For more information on this or other research topics, please visit www.aberdeen.com.

Related Research	
The Role of Data Classification in Protecting Your Intellectual Property; May 2012	*Why Don't More Enterprises Adopt Endpoint Encryption?*; March 2011
Successful IT Security Projects Invest Not Only in Technologies, But Also in People; April 2012	*Putting the P in DLP*; July 2010
Enabling Access to Big Data; April 2012	*Content-Aware: The 2010 Data Loss Prevention Report*; June 2010
Encryption, Without Tears; March 2012	*The Case for Enterprise Key Management: Higher Complexity and Scale at Lower Cost*; June 2010
Data Classification Meets Collaboration: Cross-Domain Monitoring and Filtering; February 2012	*Web Security in the Cloud*; May 2010
Does Your Enterprise Classify Its Data?; January 2012	*Email Security in the Cloud*; April 2010
Left to Their Own Devices: Does Your Enterprise Have a "Dropbox Problem"?; January 2012	*Laptop Lost or Stolen? Five Questions to Ask and Answer*; February 2010
Email and Web Security, Differentiated: Protecting Content is King; November 2011	*Enterprise Rights Management: Persistence Pays Off*; August 2009
DLP, the Ideal Referee: Let the Game Go On!; October 2011	*Microsoft SharePoint: The Comedy (and Tragedy) of the Commons*; July 2009
Secure / Managed File Transfer: Why You Should Be Looking More Closely Right Now; August 2011	*The Cost-Based Business Case for DLP*; June 2009
	Securing Unstructured Data; June 2009
	Data Loss Prevention: Little Leaks Sink the Ship; June 2008
	Aberdeen IT Security and IT GRC blogs; ongoing

Author: Derek E. Brink, Vice President and Research Fellow, IT Security and IT GRC (Derek.Brink@aberdeen.com)