FORRESTER®

For: Security &
Risk Professionals

Rethinking Data Discovery And Data Classification

by Heidi Shey and John Kindervag, October 1, 2014

KEY TAKEAWAYS

Defining Your Data Is The Foundation For Data Security And Control

Data discovery and classification is the first part of Forrester's Data Security And Control Framework, which breaks data protection into three areas: 1) defining data; 2) dissecting and analyzing data; and 3) defending data. Classification enables the creation of attributes for data identity, which helps determine how to treat and secure data.

Data Classification Does Not Have To Be Complicated

Classify new data first, and address legacy data later. Approach classification in two ways: Classify based on how the data is protected, or classify based on toxicity to determine appropriate protection. Five roles -- data creators, owners, users, auditors, and champions -- enable classification. Simplify classification levels for manageability.

Dynamic Data Classification Requires Both Tools And Human Intervention

Recognize that data is a living thing. Dynamic data classification requires the integration of both manual processes involving employees as well as tools for automation and enforcement. Human intervention provides much-needed context for data classification, while tools enable efficiency and policy enforcement.

Rethinking Data Discovery And Data Classification

Strategic Plan: The Data Security And Privacy Playbook

by Heidi Shey and John Kindervag
with Stephanie Balaouras, Cheryl McKinnon, Kelley Mak, and Claire O'Malley

WHY READ THIS REPORT

Defining data via data discovery and classification is an often overlooked, yet critical, component of data security and control. Security and risk (S&R) pros can't expect to adequately protect data if they don't have knowledge about what data exists, where it resides, its value to the organization, and who can use it. Data classification also helps to create data identity (data-ID), the missing link for creating actionable data security and control policies. Yet, S&R pros who attempt to lead efforts to classify data are thwarted by their own efforts with overly complex classification schemes and haphazard approaches. As a result, many see data discovery and classification as a Sisyphean task. This report was originally published on April 5, 2013; Forrester reviews and updates it periodically for continued relevance and accuracy, and most recently substantially revised it to factor in new ideas, tools, and data as of August 2014.

Table Of Contents

2 **Knowing Your Data Creates A Foundation For Data Security**

Data Discovery And Classification Simplify Security Controls And Policies

4 **But Data Classification Is Easier Said Than Done**

6 **Start Now, Move Forward, And Identify Data Classification Roles**

9 **Rethink Data Classification**

Approach Data Classification From Two Directions

Simplify Classification

Engage In Dynamic Classification

RECOMMENDATIONS

15 **Now That You've Defined Your Data, Put It To Work**

16 **Supplemental Material**

Notes & Resources

In developing this report, Forrester drew from insight and research through advisory and discussions with end users and vendors.

Related Research Documents

The Future Of Data Security: A Zero Trust Approach
June 5, 2014

Know Your Data To Create Actionable Policy
January 15, 2013

Rethinking DLP: Introducing The Forrester DLP Maturity Grid
January 3, 2012

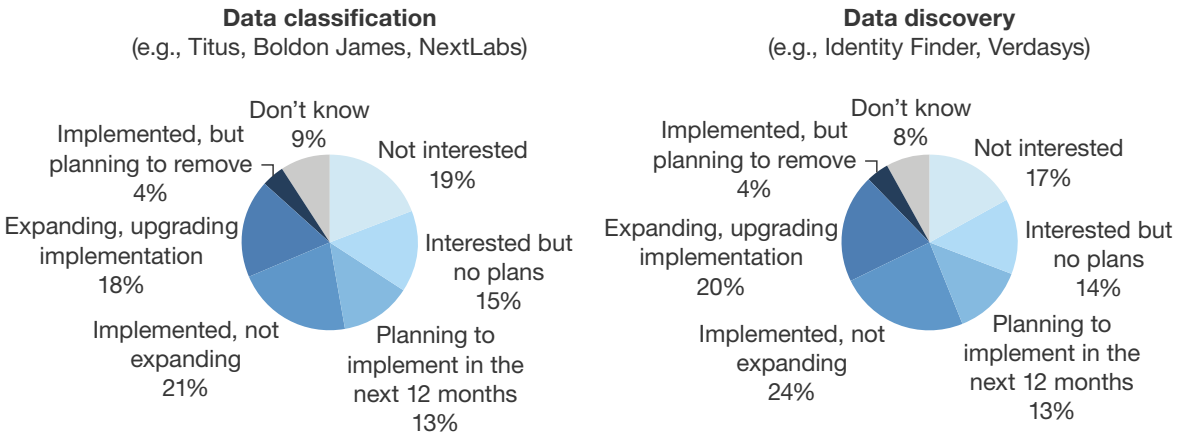
KNOWING YOUR DATA CREATES A FOUNDATION FOR DATA SECURITY

For many S&R pros, data security initiatives quickly zoom in on controlling access to data, or encrypting data. What many overlook is that understanding and knowing your data is the foundation for data security. It's silly to build a skyscraper on a foundation of sand and then be surprised when it falls down or sinks. Yet, this is how many S&R pros approach data security today. Data discovery and classification are two essential, yet often overlooked, initiatives that lay the foundation for protecting data.[1] Today, 44% of North American and European technology decision-makers at firms with 20 or more employees use data discovery tools to assist in their data security efforts, while 39% say the same for data classification tools (see Figure 1).

This foundation — defining your data — is the first part of a three-part framework called the Data Security And Control Framework that Forrester created to help S&R professionals adapt to the new data economy.[2] This framework breaks data protection into three key areas: 1) defining the data; 2) dissecting and analyzing the data; and 3) defending the data. This report is a strategy deep dive into the foundation for data security and control: defining your data.

Figure 1 Firms Are Lukewarm Toward Data Discovery And Classification Technologies

"What are your firm's plans to adopt the following data security and information risk management technologies?"



Base: North American and European technology decision-makers at firms with 20+ employees
(percentages may not total 100 due to rounding)

Source: Forrester's Business Technographics® Global Security Survey, 2014

Data Discovery And Classification Simplify Security Controls And Policies

Where's Waldo? For many firms, data is like Waldo, the ever-elusive children's book character with the red-and-white-striped shirt, bobble hat, and glasses.[3] He's hidden in elaborately drawn illustrations, filled with other characters and objects drawn to resemble him and his shirt — and it's up to the reader to find the real Waldo in the whole jumble. Imagine the pages of illustrations as your IT environment. Wally is in there; hackers know it and are doing their best to find him. Do you know where Wally — your data — is? Defining the data within the organization is a critical step: If you don't know what you have, where it is, and why you have it, you can't expect to apply the appropriate policies and controls to protect it. There are two primary functions that you must perform to help you know your data:

- **Identify the data and where it resides.** Data discovery tools and software scan endpoints or corporate network assets to identify resources that could contain sensitive information, such as hosts, database columns and rows, web applications, storage networks, and file shares.

  Data discovery tools and software from a variety of vendors such as Digital Guardian (the vendor formerly known as Verdasys), Ground Labs, Identity Finder, and StoredIQ (an IBM company) help enterprises identify the locations of sensitive structured and unstructured information. Their primary goal is to find assets that you can then classify. S&R pros may want to consider using such tools to help automate the discovery process. Manual discovery is a time-consuming and often error-prone process. Data discovery tools and software are distinct from, but related to, data classifiers. They are also increasingly included in data loss prevention (DLP) suites such as those from McAfee, RSA, Symantec (Vontu), and Websense. eDiscovery and information governance tools such as AccessData, Active Navigation, Guidance Software, Nuix, Recommind, and Zylab can also help to find data assets.[4]

- **Apply labels to classify the data and determine how it's handled.** Data classification tools generally look for data that they can match deterministically, such as credit card numbers or Social Security numbers (SSNs). Some data classification tools also use fuzzy logic, syntactic analysis, and other techniques to classify less-structured information.

  Data classification tools for security parse structured and unstructured data, looking for sensitive data that matches predefined patterns or custom policies established by customers. Once the data classification tools have identified matches, they apply labels to the information so that it may be protected by, for example, DLP tools.[5] When this sensitive data is present in documents — i.e., SSN in a PDF or Microsoft Word doc — many enterprises may also remediate by moving the data into a content management or archive system that can then apply access controls or metadata labels to restrict usage. Many data classification tools today also support user-driven classification to engage workers who can provide context about the data and its sensitivity level. Examples of vendors that offer solutions to help with data classification include Boldon James, Concept Searching, Digital Guardian, Imperva, NextLabs, Titus, Varonis Systems, and Whitebox Security.

While it's important to recognize that data classification is integral to an effective DLP initiative, it's not simply a DLP-related effort or a form of DLP. Classification is the foundation for all data security, including DLP. Without data classification in play, it's impossible to know what data to protect. Your data classification will provide enormous value by enabling the creation of attributes for data. Adding these attributes to data gives it an identity. This data-ID (D-ID) created from these attributes and metadata tags to data packets serves to help both technology and people make decisions on what to do with this piece of data and how to handle it appropriately.[6] In addition, data classification aids in other security activities such as monitoring and access control reviews; it can also help realign focus and costs by protecting valuable data while allowing unclassified (public) data to live in a less monitored environment.

BUT DATA CLASSIFICATION IS EASIER SAID THAN DONE

There are only two types of data that exist in your organization: 1) data that someone wants to steal and 2) everything else. The first type is sensitive or toxic data, which you can easily identify with the equation 3P + IP = TD. The three P's stand for personal cardholder information (PCI), personal health information (PHI), and personally identifiable information (PII); IP is intellectual property; and TD is toxic data (see Figure 2). S&R pros must discover and help rally the organization to classify toxic data. In general, when aided by data discovery tools, data discovery is a more straightforward process than data classification. Even with tools in place, data classification can be a messy endeavor when:

- **Awareness and importance of classification are lacking.** Let's face it: Data classification is not the sexiest business or IT initiative out there. The conceptual argument for why it's important is easy enough to grasp, but practical implementation requires a great deal of management support and focus. Without this support, data classification becomes an academic feel-good exercise, resulting in a policy that collects dust on a shelf or merely an ad hoc and spotty implementation within the organization.

- **Classification schemes and terms are overly complicated.** There's an inclination to over-classify data. A typical enterprise data classification scheme has anywhere from three to six levels (see Figure 3). Forrester has seen organizations with as many as nine levels of classification. The complexity involved with identifying unique criteria for so many levels — and then having employees understand the difference and apply labels correctly — is a ticking time bomb waiting to explode. Other dimensions such as likelihood and impact of a data breach may be taken into consideration. The terms used to describe classification levels may not make it immediately obvious what the classification really means. And in organizations that do business with government clients, the language and terms used for classification labels can cause confusion if they differ from those within a government setting.

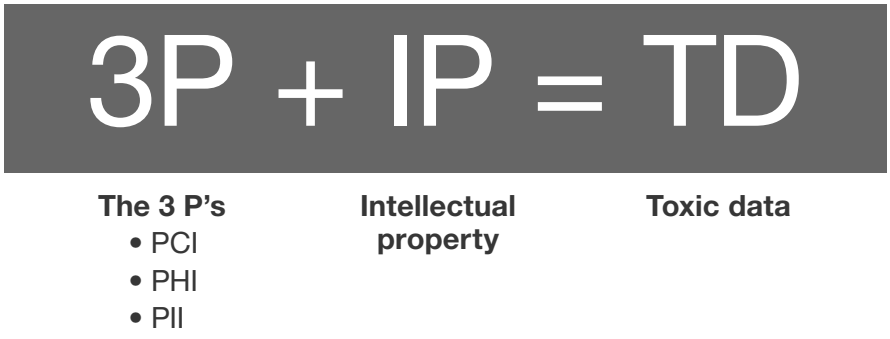
- **Classification schemes are outdated and unrealistic.** For the most part, data classification is a subjective exercise. Two people can look at the same piece of data and give very different classification levels for it, especially when a complex classification scheme is in use. Organizations with a long-established classification scheme — for instance, one that's been around for 20 years — may have trouble making changes or hesitate to make changes despite the loopholes that may be apparent with the legacy classification scheme. In many of these situations, classifications are simply not usable or realistic for enforcement in the current technology environment. Another challenge that makes classification unrealistic is when the classification scheme itself does not align with — or actually conflicts with — other security and data use policies in place within the organization.

- **A global workforce adds additional complexity to classification efforts.** Rolling out an overarching data classification scheme for a large, multinational organization without a clear understanding of local or regional data-handling considerations is the first step toward a stalled or failed implementation. There are a multitude of data privacy-related legal obligations to consider, as well as insight from various internal stakeholders who hold data privacy responsibilities.[7] In addition to the data considerations, there are also language considerations when applying labels to data. This is where visual markings or tags can help create a unified understanding of label terms.

- **Roles and responsibilities are unclear.** Each organization has its own cultural history of data responsibility and data ownership. Various groups will also have different perceptions of the value of data and information, which colors their view of appropriate data-handling and data security needs. This makes for opportunistic classification, or classifying data in such a way that makes it usable for that group's needs rather than classifying data in a way that is risk-appropriate for the organization. A lack of checks and balances with data roles and responsibilities exacerbates the issue of opportunistic classification.

- **Data classification has different meanings within the organization.** Different groups may view classification differently, and not from a security perspective. For example, an individual overseeing information knowledge management could have a data classification initiative that involves categorization of data (identifying if a data file is a contract versus a memo) rather than sensitivity of data. S&R pros must understand the rationale behind the different types of classification that may exist within the organization, and work together with other classification stakeholders to combine efforts and present a cohesive strategy and business case for classification to the enterprise.

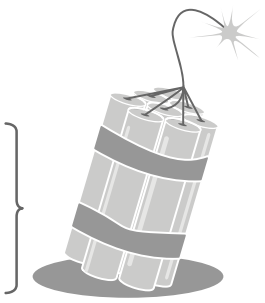
Figure 2 Toxic Data Consists Of Sensitive Information



85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

Figure 3 A Typical Information Classification Scheme

#	Level	Description
1	Public	Anything not company-internal
2	Internal	Internal but not for public release (e.g., earnings)
3	Confidential	Not for distribution (e.g., memos, plans, strategy)
Additional classification levels as appropriate . . .		
X	Restricted	Highly compartmentalized (e.g., salaries, regulatory information, SOX-controlled)



85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

START NOW, MOVE FORWARD, AND IDENTIFY DATA CLASSIFICATION ROLES

The most difficult part of data classification is getting started. Recognize that there are two types of data classification projects: new data and legacy data (see Figure 4). Commit to starting data classification for new data first for immediate gains. Data classification is not one person's job. It's everyone's job. Clearly define data classification roles and responsibilities within your organization to embed data classification processes into normal business processes (see Figure 5):

- **Data creators.** Anyone within the organization can be a data creator. The responsibility of identifying this new, freshly created piece of data as toxic or nontoxic rests with its creator. Data creators can ask themselves one simple question to determine toxicity: Would it be acceptable for this data to find its way into a competitor's hands? If not, it's toxic data. While they are the source of this new data within the organization, data creators should spend the least amount of time and effort classifying data, and get back to doing their jobs.

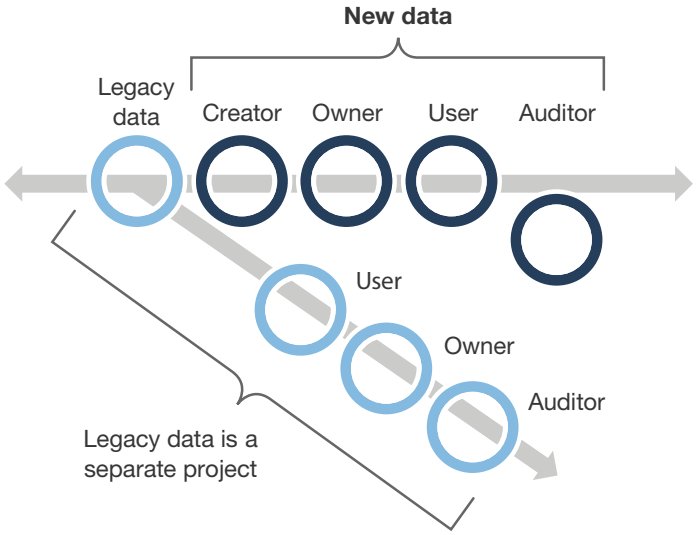
- **Data owners.** A data owner may be a line of business manager, division head, or equivalent. If the data resides and is primarily in use within their group, they own it. The data owner must review the data creator's toxic/nontoxic label. If there is agreement, determine the classification level for the data based on how the data can be used. If there is disagreement with the toxicity assessment, start a discussion as to why, before applying a classification label. Data owners are also likely to be privy to any data use agreements between the firm and its third-party business partners. The purpose here is for the data owner to provide context in classification, which automated tools lack.

- **Data users.** Anyone who has access to this data is a data user. Those who are authorized to handle and use the data are in the best position to provide feedback about the data classification tags: Is the classification appropriate based on how the data is used? Are there circumstances or situations where the data is used or could be handled differently than what's allowed under the current classification?

- **Data auditors.** A data auditor may be a risk and compliance manager, a privacy officer, a data officer, or equivalent role. The data auditor must review the data owner's assessment of classification and determine if it's in line with business partner, regulatory, or other corporate requirements. The data auditor also reviews feedback from data users and assesses alignment between actual or desired data use and current data-handling policies and procedures.

- **Data champion.** A data champion is an individual responsible for the organization's use of data for business purposes, and thus has an incentive to ensure that the data is protected and used appropriately. This role can emerge in different forms. A chief data officer responsible for data strategy, including quality, governance, and monetization, may take on this duty.[8] A data steward responsible for data governance, practices, and requirements may be tapped to fill this role too.[9] The key here is to ensure that there is an identified business stakeholder who will support and drive data classification efforts as a part of the organization's overall data strategy.

Figure 4 There Are Two Types Of Data Classification Projects



85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

Figure 5 Data Classification Roles

	Data creator	Data owner	Data user	Data auditor	Data champion
Who this is	All employees	Line of business manager	All employees	Risk/compliance officer, chief privacy officer, or equivalent	Chief data officer, data steward, or equivalent
His or her role in data classification	Decides if data is toxic or nontoxic at time of creation	• Reviews data creator's toxic or nontoxic label • Determines the classification level for the data	• Uses the data • Provides feedback on data classification level	• Reviews data owner's assessment of classification level for the data • Reviews data user feedback to assess alignment between actual or desired data use	• Is responsible for data strategy • Oversees and provides support for data classification efforts

85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

RETHINK DATA CLASSIFICATION

There are some common approaches to data classification that complicate the process and reduce it to an academic exercise rather than a practical implementation. Rethink data classification from four key angles (see Figure 6):

- **Rethink: New data classification.** The traditional solution is to apply classification labels to new unstructured content. Labels, tags, and markings are important, but they are not the root cause of the problem. The issue with data classification for new data is enforcement and how to address changes in classification when the need arises. Plan for feedback loops and points in time where data can be reclassified as needed. Rely on access control and enterprise rights management for the most sensitive data (radioactive data). Turn to DLP, access controls, content management systems, and best effort for all other types of data.

- **Rethink: Educating employees about classification.** The traditional solution in a data classification project rollout is to educate employees about the different classification levels, their respective markings, and when to apply them. The challenge here is that if these levels are not clear-cut and easily discernible, data classification becomes subjective and opportunistic. Employees don't have time — or much desire — to parse through multiple levels of classification labels; they just want to do their job. Re-engineer the workplace so that this type of thinking is not required. The task of applying a data classification label needs to be as simple a decision as possible. Forrester recommends no more than three classification levels.

- **Rethink: Data sets and data.** They are not the same. A Social Security number is a piece of data. A data set is a collection of data, where some of it may be toxic or sensitive and some of it not. With a data set, the output of data mining or a report with combined data elements that are a mix of nontoxic and toxic individually can become toxic in aggregate. Focus on identifying the toxicity of data, not data sets, to address this concept of toxic transference. For example, the date of a purchase order or SKU of an item that was purchased is not toxic data, but can be included in a data set that does include toxic data like credit card numbers. If a piece of data is toxic on its own (e.g., credit card number, shipping address), it will transfer its toxicity when combined and used with other nonsensitive data (e.g., date of purchase order, item purchased).

- **Rethink: Legacy data classification.** The task of classifying existing, legacy data is a separate project. Obsessing over legacy data inhibits an organization's ability to start a data classification project. There is so much data that the task seems insurmountable. Don't worry about the old data until you've established a process for classifying new data. Once that's in place, classifying legacy data will not seem as daunting. Yet, be realistic and recognize that a legacy data classification project can be time-consuming and error-prone, especially if data creators have left the company or those who use the data today are handling it in ways that the data was not originally intended for. Often, questions of whether or not this data should even be kept begin to surface — and this is a good thing! Don't hang on to excess data if you don't have to, because those data assets can morph into a liability.

S&R pros can choose to tackle legacy data classification through a brute-force search or leverage an eDiscovery solution.[10] Implement or update access controls upon discovery of this legacy data. Force the declassification of this data after three years' time, or implement a review for reclassification if necessary. In most cases, by this age, the value and sensitivity of internal data has diminished to the point that the data can either be disposed of or kept but considered unclassified (public) information.

Figure 6 Rethink Information Classification

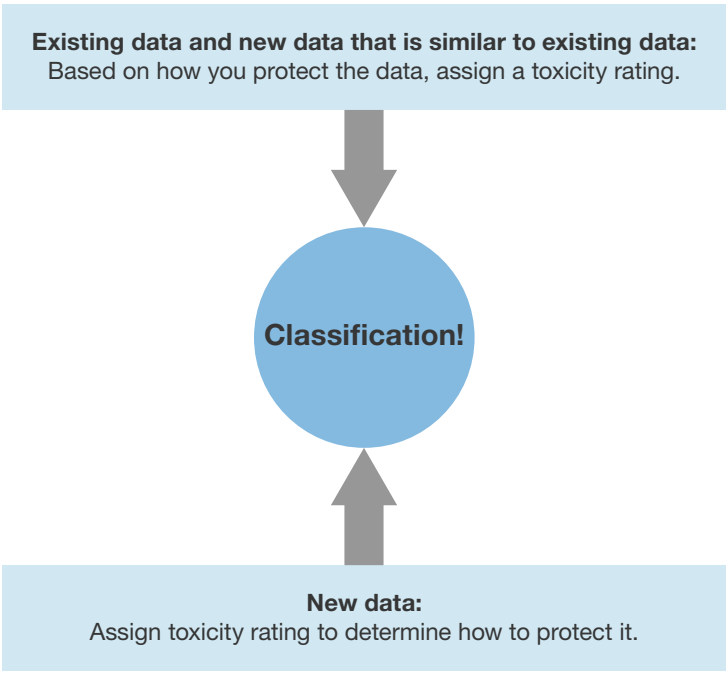
Traditional solution	Reality	Practical solution
Classify existing information assets.	• Time-consuming and error-prone • No labeling support	• Brute-force search • eDiscovery • Access control • Forced declassification after three years
Educate employees.	Employees don't have time to think.	• Re-engineer workplace so thinking isn't required. • Three classification levels max
Treat data sets and data the same.	• They are not the same. • Data toxicity is transferrable.	Focus on identifying and classifying data, not data sets.
Apply classification labels to new unstructured content.	No universal technology support	• Dynamic classification from content • Access control and ERM for radioactive (highest level) information • DLP, access controls, best effort for everything else

85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

Approach Data Classification From Two Directions

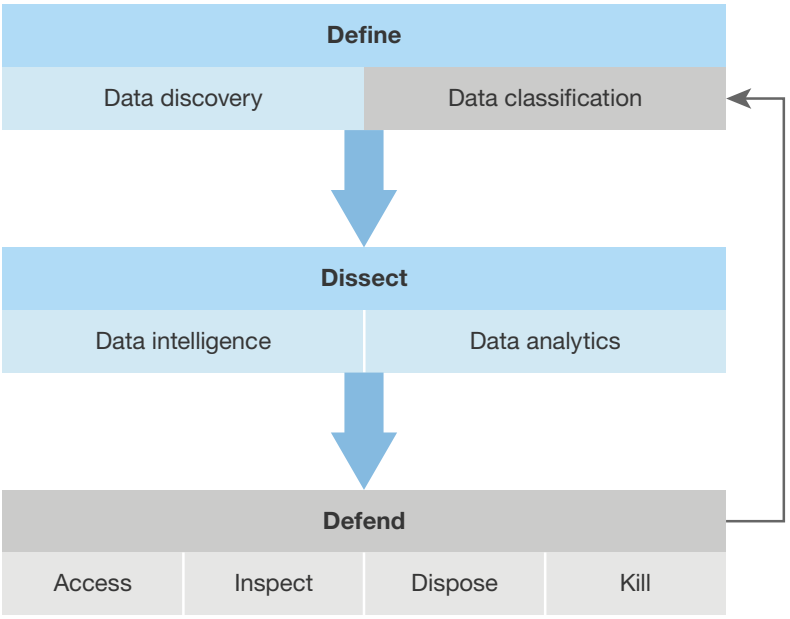
While it's possible to determine the appropriate protection needed for data based on how it's classified, S&R pros can also classify data based on how it's protected (see Figure 7). Classifying data based on the data creator or user's assessment encourages a deeper understanding of how the data is used and valued, providing guidance as to how it should be protected. Classifying data based on how it's protected encourages a better understanding of existing data and how it should be classified based on use. When we consider the protections applied to data within the Data Security and Control Framework, there are really only four levers to pull: access controls, inspection of usage patterns, data disposal, and killing data (devaluing data through encryption, tokenization, or data-masking technologies) (see Figure 8).[11]

Figure 7 Two Approaches For Data Classification



85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

Figure 8 How Data Is Protected Can Inform Classification



85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

Simplify Classification

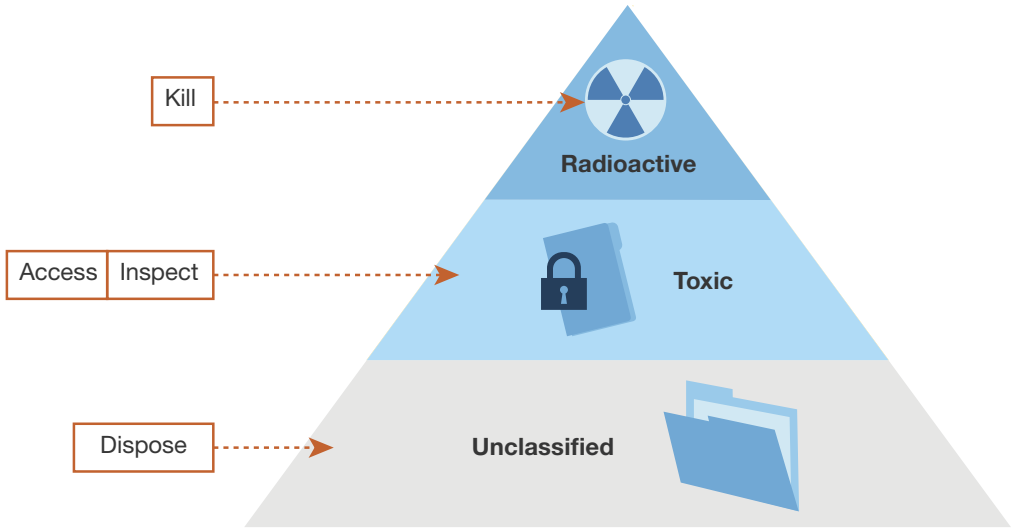
S&R pros can become too entrenched in discussions about the differences between structured and unstructured data, creating a veil of complexity. Data is data, and its classification can be simplified into three tiers: radioactive, toxic, and unclassified (see Figure 9):

- **Radioactive.** Radioactive data consists of toxic data: PCI, PHI, PII, and IP. It's data that is subject to local or national laws or compliance regulations. Data is also radioactive when its loss will violate a business agreement. You should protect radioactive data primarily through robust technical controls. If encryption, tokenization, or data-masking technologies must be used to protect the data, the data is considered radioactive.

- **Toxic.** Data is toxic when its loss will do harm to customers or employees and likely incur significant costs for the firm and cause brand damage. It may on rare occasions also consist of IP (which is much more likely to be considered radioactive data). You should protect toxic data primarily through policy and procedures. If strict access controls and inspection of data usage for suspicious or anomalous behaviors are used to protect the data, the data is considered toxic.

- **Unclassified.** Data is unclassified when one can treat and handle it as public information without harm to the organization and its employees or customers. Over time, the organization can likely reclassify internal data as unclassified data as its sensitivity level diminishes. For example, within a public company, earnings information may be internal data up until the time of the company's earnings call with investors. At that point, the information is unclassified. If disposal of the data is acceptable, the data is considered unclassified or public.

Figure 9 Simplify Data Classification With Three Levels



85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

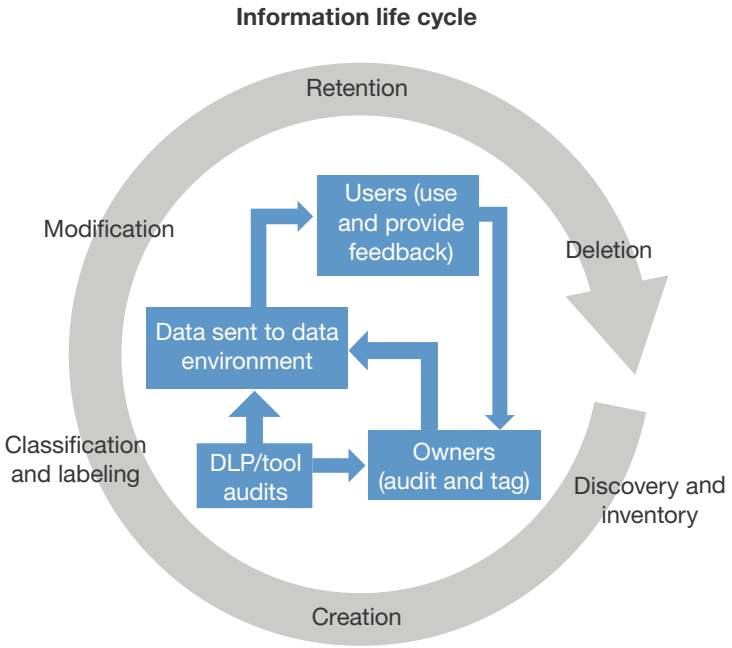
Engage In Dynamic Classification

Treat data as living, not static. Its value is highest at the point of creation, and over time may diminish. Over the course of a piece of data's life cycle, classification should be continuous (see Figure 10). Classification is a dynamic and circular process that involves both manual and automated processes. It is composed of two feedback loops that make it dynamic (see Figure 11). One method of classification involves identifying the data types that exist in the organization, and the protections that currently exist for that data in order to determine a classification level for that data (see Figure 12).

Another method of classification starts at the point of data creation. An example workflow for this also aligns with the three classification levels (see Figure 13). At the time of creation, the data creator tags the data as toxic or nontoxic and sends it off to the data environment (a location where the data is stored and used). Data users then proceed to use this data as they normally would. Data owners validate this tagging, and data users provide feedback as to whether the classification is correct. Through these actions, data owners and data users provide an audit mechanism for data classification. This part of the workflow and feedback loop is largely a manual process aided by tools, providing much-needed context and identification information for data.

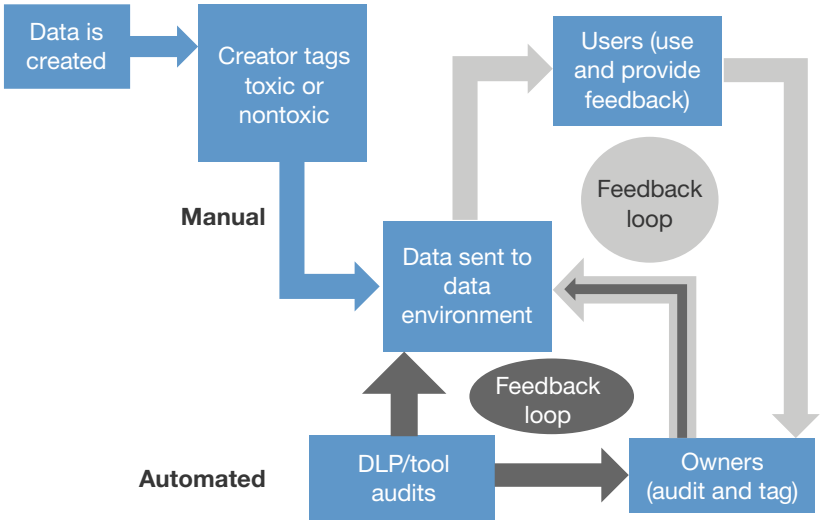
Automation also has a place in dynamic data classification. It automates another feedback and audit loop between tools like DLP, the data, and classification tags that data owners put in place. You must involve data owners with the implementation of any tool for automation to help feed and tune these tools with the necessary information to build this system.

Figure 10 Dynamic Data Classification Is A Key Part Of The Information Life Cycle



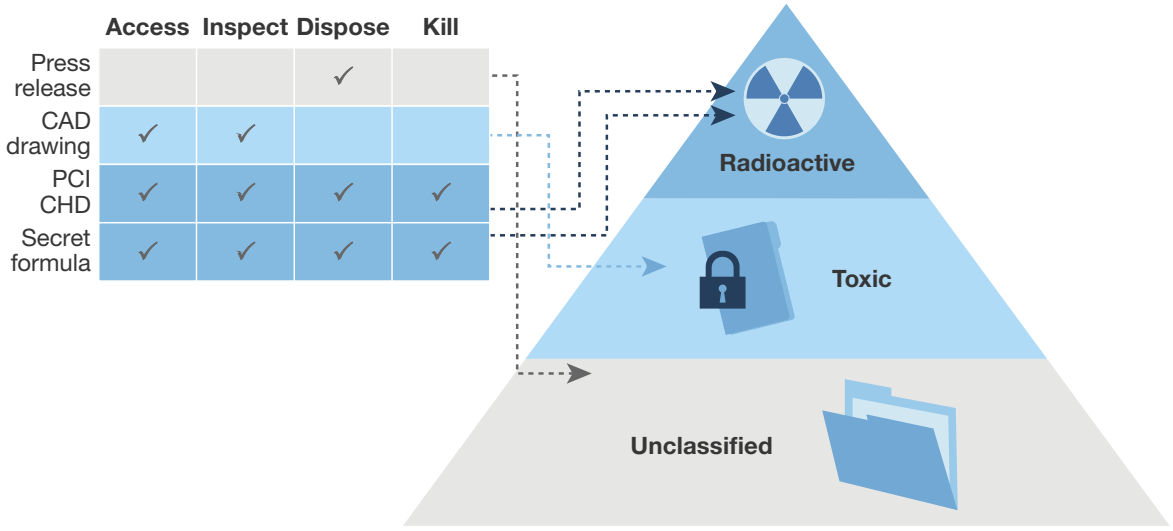
85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

Figure 11 Dynamic Data Classification Involves Two Feedback Loops



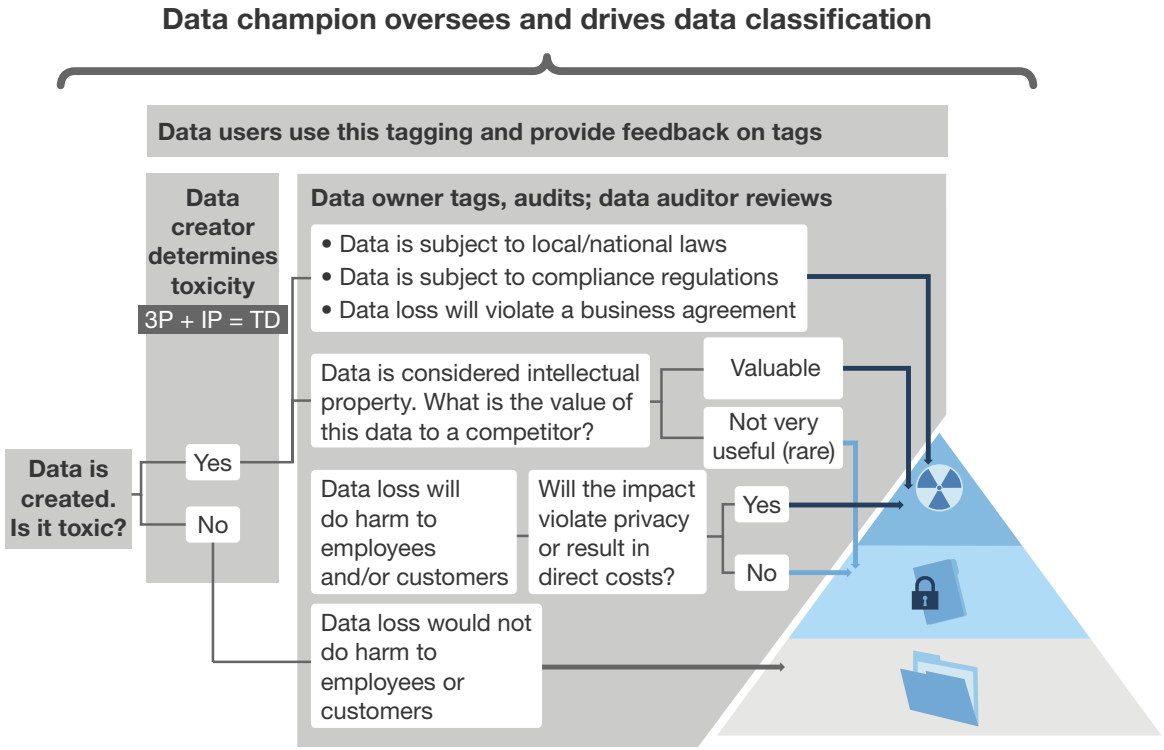
85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

Figure 12 An Example Of Defining Data Based On How It Is Protected



85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

Figure 13 An Example Workflow For Defining Data From The Point Of Creation



85842 Source: Forrester Research, Inc. Unauthorized reproduction or distribution prohibited.

RECOMMENDATIONS

NOW THAT YOU'VE DEFINED YOUR DATA, PUT IT TO WORK

Dynamic data classification is not the end, but a beginning, for data defense. As part of ongoing data defense, we recommend that S&R pros:

- **Consolidate your data.** After engaging in data discovery and classification — defining your data — S&R pros usually find that data exists in unlikely or unexpected places. This is a great opportunity to then consolidate the data, aggregating it into fewer places to limit breach exposure and potentially reduce the scope of compliance for mandates like PCI. Data consolidation is also a critical process stage in DLP maturity.[12]

- **Create or revisit policies for data use and protection.** Knowing and understanding the data allows for actionable data security policies. The next step then is to understand the implications for data-handling from storage to disposal, and consider audit mechanisms for

policy enforcement such as DLP and network analysis and visibility (NAV) tools. Specific data security policies to address include access control, data inspection and usage, data disposal, and data encryption.[13]

- **Determine when you should expire data.** The concept of dynamic data classification treats data as a living thing that is able to — and should — change when the need arises. Changes will occur within the business, and the classification level of data will need to align accordingly. In the case of a merger or acquisition, merger data is extremely sensitive within the time frame leading up to the merger, but mostly public after the merger. Determine how and when to expire data in line with approved corporate retention policies as defined by legal, compliance, or information management peers.

SUPPLEMENTAL MATERIAL

Survey Methodology

For the Business Technographics® Global Security Survey, 2014, Forrester conducted a mixed methodology phone and online survey fielded in April and May 2014 of 3,305 business and technology decision-makers located in Australia, Brazil, Canada, China, France, Germany, India, New Zealand, the UK and the US from companies with two or more employees.

ENDNOTES

[1] Forrester has created a framework to help security and risk professionals control big data. We break the problem of securing and controlling big data down into three areas: 1) defining the data; 2) dissecting and analyzing the data; and 3) defending and protecting the data. See the June 5, 2014, "The Future Of Data Security: A Zero Trust Approach" report.

[2] S&R pros must control and protect the extreme volumes of data that your organization aggregates in a big data environment. And ideally, now is the time to bring together separate silos of data control and protection such as archiving, DLP, and access control. This also involves moving data security controls closer to the data itself, instead of at the edges (perimeters) of networks. Forrester has created a framework to help security and risk professionals control big data. We break the problem of securing and controlling big data down into three areas: 1) defining the data; 2) dissecting and analyzing the data; and 3) defending and protecting the data. See the June 5, 2014, "The Future Of Data Security: A Zero Trust Approach" report.

[3] Waldo is also known as Wally, Holger, Charlie, Hetti, Walter, Effy, or Hugo, depending on the country in which his book is published. Waldo travels to everyday places — such as the beach, the ski slopes, and the zoo. The original book from 1987 features 12 detailed two-page illustrated spreads of the different locations. Somewhere amid the crowded scene is Waldo, and readers are asked to scour the detailed illustrations to locate the lost traveler. Source: Where's Waldo (http://whereswaldo.com/index.html#home).

[4] Organizations are insourcing more of the eDiscovery process as information management programs mature and as technologies expand to meet the needs of corporate legal teams — not just law firms and legal service providers. This report covers 37 eDiscovery solutions for IT, legal, and records management professionals to consider. See the December 13, 2013, "Market Overview: eDiscovery, Q4 2013" report.

[5] Tools for data discovery, classification, and DLP currently have overlapping functionality. Long term, we expect to see greater convergence of data classification and DLP tools. See the April 22, 2014, "Tech Radar™: Data Security, Q2 2014" report.

[6] Data identity is a key concept for actionable data security policy. Applying identity and tagging data packets with identity attributes allows us to determine the business criticality of any piece of data, and thereby protect it more effectively. Data identification must address three things: data identity, data handling roles, and data control tools. See the January 15, 2013, "Know Your Data To Create Actionable Policy" report.

[7] Privacy is like other critical functions within the organization. It is an ongoing process, not a one-time planning or triggered event. While securing or protecting an individual's personally identifiable information (PII) from unauthorized use or theft is critical, it's just one aspect of privacy. This is why it's critical that you work with groups or departments from legal to HR to address it. As both privacy laws and data volumes explode, it's becoming an increasingly difficult task to both comply with regulations and prevent privacy infringements, while supporting business innovation and expansion. Chief privacy officers will weigh in not only on data security and privacy issues relating to customer data but also on corporate and employee data. See the September 12, 2012, "Identify And Influence Data Security And Privacy Stakeholders" report and see the August 23, 2012, "Job Description: Chief Privacy Officer" report.

[8] The CDO role is still a relatively young one across many organizations, and much like the emerging role of CIO 20 years ago, the core responsibilities of this position are not yet fully established. Research on the role of the CDO has shown that today's CDOs are primarily responsible for data strategy, including data governance, data infrastructure, enterprise analytics, and enterprise data asset development. See the January 15, 2013, "Know Your Data To Create Actionable Policy" report.

[9] In anticipation of the increasing adoption of personal identity and data and data management (PIDM) tools and services, customer insights (CI) leaders will be held increasingly accountable for their organizations' data collection, management, and use practices — including those of the vendors they hire to augment their CI teams. This practice, which Forrester calls "data stewardship," is an imperative that organizations must plan for and enact today. See the February 22, 2013, "Building Data Stewardship Is A New Customer Insights Imperative" report.

[10] This report covers 37 eDiscovery solutions for IT, legal, and records management professionals to consider. See the December 13, 2013, "Market Overview: eDiscovery, Q4 2013" report.

[11] Access, inspect, dispose, and kill are the four components of the Defend phase of the Data Security And Control Framework. See the June 5, 2014, "The Future Of Data Security: A Zero Trust Approach" report.

[12] It can be difficult to tell your DLP tool what data to look for, alert on, or block. To help our customers characterize a more effective DLP process, Forrester has defined five process stages of DLP maturity: discover, classify, consolidate, design, enforce. See the January 3, 2012, "Rethinking DLP: Introducing The Forrester DLP Maturity Grid" report.

[13] Too often, organizations create data policies without a clear understanding of feasibility and purpose within their business because they themselves are in the dark about their data — from what data they have to where it resides. As a result, many data security policies are ineffective and can even hinder business processes. To help security professionals adapt to the new data economy, Forrester has created our Data Security And Control Framework. This framework breaks data protection into three key areas: 1) defining the data; 2) dissecting and analyzing the data; and 3) defending the data. Security pros can build a policy layer on top of this control framework where: 1) defining the data leads to identifying the data; 2) dissecting and analyzing the data leads to understanding data implications and creating audit mechanisms; and 3) defending and protecting the data leads to creating data security and control policies. See the January 15, 2013, "Know Your Data To Create Actionable Policy" report.

About Forrester

A global research and advisory firm, Forrester inspires leaders, informs better decisions, and helps the world's top companies turn the complexity of change into business advantage. Our research-based insight and objective advice enable IT professionals to lead more successfully within IT and extend their impact beyond the traditional IT organization. Tailored to your individual role, our resources allow you to focus on important business issues — margin, speed, growth — first, technology second.

FOR MORE INFORMATION

To find out how Forrester Research can help you be successful every day, please contact the office nearest you, or visit us at www.forrester.com. For a complete list of worldwide locations, visit www.forrester.com/about.

CLIENT SUPPORT

For information on hard-copy or electronic reprints, please contact Client Support at +1 866.367.7378, +1 617.613.5730, or clientsupport@forrester.com. We offer quantity discounts and special pricing for academic and nonprofit institutions.

Forrester Focuses On Security & Risk Professionals

To help your firm capitalize on new business opportunities safely, you must ensure proper governance oversight to manage risk while optimizing security processes and technologies for future flexibility. Forrester's subject-matter expertise and deep understanding of your role will help you create forward-thinking strategies; weigh opportunity against risk; justify decisions; and optimize your individual, team, and corporate performance.

« **SEAN RHODES,** client persona representing Security & Risk Professionals