Rethinking Data Discovery And Data Classification Strategies

Strategic Plan: The Data Security And Privacy Playbook

by Heidi Shey and John Kindervag

March 25, 2016

Why Read This Report

Defining data via data discovery and classification is an often overlooked, but critical, component of data security and privacy. 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, how valuable it is to the firm, and who can use it. In this report, we help S&R pros rethink overly complex and haphazard legacy approaches to discovery and classification. With the right approach, S&R pros can craft policies and deploy the right mix of controls that will protect customer data and the firm's intellectual property.

This is an update of a previously published report. Forrester reviews and revises it periodically for continued relevance and accuracy and most recently did so to factor in new ideas, tools, and data.

Key Takeaways

Defining Your Data Is The Foundation For Data Security And Privacy

Data discovery and classification is the first part of Forrester's Data Security And Control Framework, which breaks down 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 by how you protect the data or by how toxic it is 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 Strategies

Strategic Plan: The Data Security And Privacy Playbook

by Heidi Shey and John Kindervag
with Stephanie Balaouras, Cheryl McKinnon, Kelley Mak, Alexander Spiliotes, and Kara Hartig
March 25, 2016

Table Of Contents

Notes & Resources

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

Related Research Documents

The Future Of Data Security And Privacy: Growth And Competitive Differentiation

Know Your Data To Create Actionable Policy

Rethinking DLP: Introducing The Forrester DLP Maturity Grid

FORRESTER®

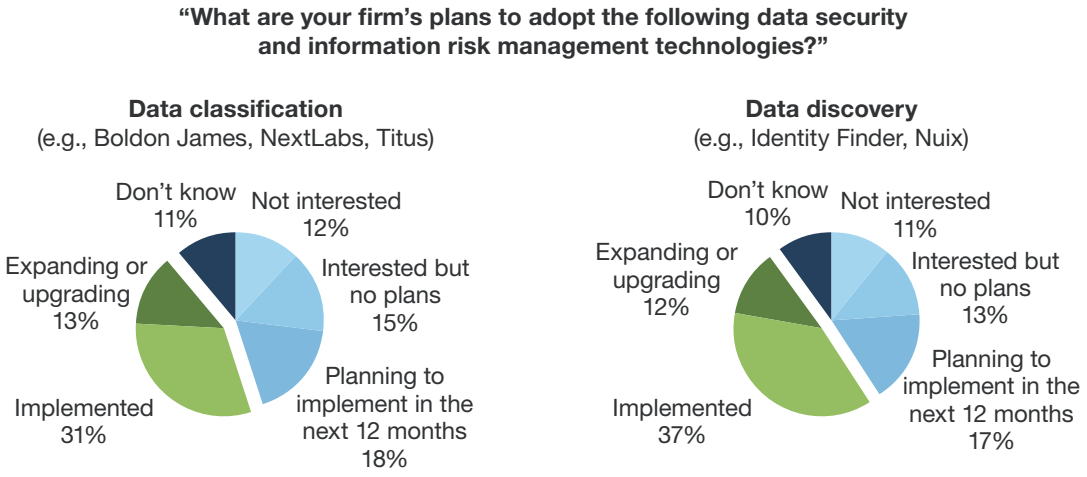
Forrester Research, Inc., 60 Acorn Park Drive, Cambridge, MA 02140 USA
+1 617-613-6000 | Fax: +1 617-613-5000 | forrester.com

Knowing Your Data Creates A Foundation For Data Security And Privacy

For many S&R pros, data security initiatives quickly zoom in on controlling access to data or encrypting data. But many overlook that understanding and knowing your data is the foundation for both data security and privacy. Today, 49% of North American and European enterprise software decision-makers use data discovery tools to assist in their data security efforts, while 44% 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. Forrester created it to help S&R professionals adapt to the 1) evolving cyberthreats to sensitive data; 2) increasingly complex regulatory landscape; and 3) customers' changing attitudes and behaviors regarding privacy.[1] This framework breaks data protection down 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 More Are Turning To Data Discovery And Classification Technologies



Base: 436 North American and European technology decision-makers (in firms with 1,000+ employees)
(percentages may not total 100 due to rounding)

Source: Forrester's Global Business Technographics® Security Survey, 2015

Data Discovery And Classification Simplify Security Controls And Policies

To respect privacy and meet data protection requirements, you must understand what constitutes personal information, how the firm collected this data, why it collected this data, and how it uses it —while ensuring that you can enforce your privacy policy. You also can't protect your intellectual property from cybercriminals and agents of corporate espionage if you don't know its value to the firm (and its value in cyber black markets and to competitors). That's why defining the data within the firm 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. 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.

  Available solutions: Data discovery tools and software from a variety of vendors such as Digital Guardian, Ground Labs, Identity Finder, and StoredIQ (an IBM company) help enterprises identify the locations of sensitive structured and unstructured information. eDiscovery and information governance tools such as AccessData, Active Navigation, Guidance Software, Nuix, Recommind, and Zylab can also help find data assets.[2] Data discovery capabilities are also increasingly included in data loss prevention (DLP) suites or platforms such as those from Clearswift, Digital Guardian, Forcepoint (formerly Websense/Raytheon), McAfee, and Symantec.

- **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 that customers establish. Once the data classification tools have identified matches, they apply labels to the information so that, for example, DLP tools can protect it.[3]

  Available solutions: When this sensitive data is present in documents (e.g., an 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, and Varonis Systems.

Data Classification Gives Data An Identity And Builds On Existing Security Initiatives

Recognizing that data classification is integral to an effective DLP initiative is important, but it's not simply a DLP-related effort or a form of DLP. Classification is the foundation for all data security and privacy-related efforts, including DLP. This is because:

- **Without classification, intellectual property that you should protect is left vulnerable.** Some types of sensitive data like credit card numbers or account numbers are immediately obvious, and machines easily detect them. Other types like trade secrets or intellectual property aren't always obvious. Your data classification will provide enormous value by enabling the creation of attributes for data. Adding these attributes to data gives it an identity. The data-ID (D-ID) that results from these attributes and metadata tags to data packets serves to help both technology and people make decisions on what to do with a piece of data and how to handle it appropriately.[4]
- **With classification, you not only build a foundation but also optimize security efforts.** Other security activities such as monitoring and access control reviews as well as employee security awareness benefit from data classification. Classification 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

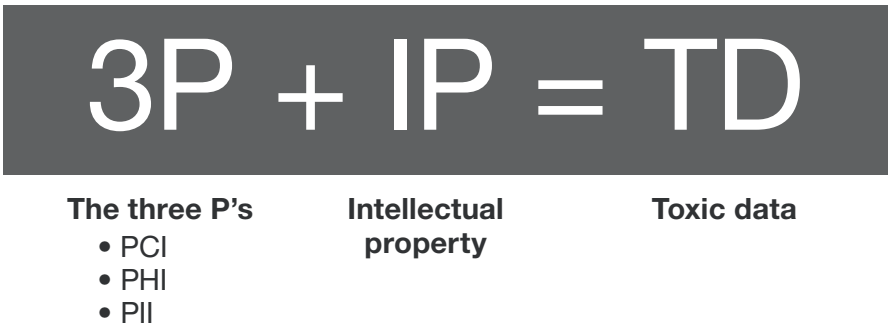
There are only two types of data that exist in your organization: 1) data that someone wants to steal and 2) everything else. The first type is sensitive or toxic data, which you can easily identify with the equation 3P + IP = TD. The three P's stand for personal cardholder information (PCI), personal 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, with the aid of 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 initiative. 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. Several firms that we spoke to were in the process of a classification initiative reboot for this reason; they had a policy in place, but most of the company didn't follow it, rendering it ineffective for business and security purposes.
- **Classification schemes and terms are overly complicated.** There's an inclination to overclassify data. Forrester has seen firms with as many as nine levels of classification. The complexity of identifying unique criteria for so many levels — and then having employees understand the difference and apply labels correctly — creates unnecessary hurdles.[5] The terms for describing

classification levels may also obscure what the classification levels really mean. This contributes to opportunistic classification — in other words, classifying data to make it usable for a group's needs rather than classifying data so that it is risk-appropriate for the organization.

- **Classification schemes are outdated and unrealistic.** 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 exist in the legacy classification scheme. In many of these situations, classifications are simply not usable or realistic for enforcement in the current technology environment.
- **A global workforce and unclear responsibilities add complexity to classification.** Each firm has its own cultural history of data responsibility and ownership. Different perceptions of the value of data can exist across different groups, which color their view of appropriate data-handling and security needs. Rolling out an overarching data classification scheme for a large multinational firm without a clear understanding of corporate culture and local or regional data-handling considerations is the first step toward a stalled implementation.[6] Gather insight from various internal stakeholders who hold data privacy responsibilities.[7] There are also language considerations when applying labels to data. This is where visual markings or tags can help create a unified understanding of labels.
- **Data classification has different meanings within the organization.** Different groups may view classification differently, and not from a security and privacy perspective. For example, an individual overseeing information knowledge management could have a data classification initiative that involves the categorization of data (e.g., identifying if a data file is a purchase order versus a memo) rather than the sensitivity of data. S&R pros must understand the rationale behind the different types of classification that may exist within the organization. They must then work together with other classification stakeholders to combine efforts and present a cohesive strategy and business case for classification to the enterprise.[8]

FIGURE 2 Toxic Data Consists Of Sensitive Information



Rethink Your Data Classification Strategy

There are some common approaches to data classification that complicate the process and reduce it to an academic exercise rather than a practical implementation. To combat this, rethink data classification from four key angles (see Figure 3):

- **New data classification and the data life cycle.** The traditional solution is to apply classification labels to new unstructured content and call it a day. The issue with data classification for new data is enforcing it and adjusting the classification when the need arises. Plan for feedback loops and points in time when you can reclassify data if necessary. Rely on access control and enterprise rights management (ERM) for the most sensitive data (i.e., radioactive data). Turn to DLP, access controls, content management systems, and best efforts for all other types of data.
- **Education for 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 the application of them. The challenge here is that if these levels are not 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. The task of applying a data classification label needs to be as simple as possible. Forrester recommends no more than three classification levels.
- **Data sets and data.** They are not the same. An SSN is a piece of data. A data set is a collection of data, typically in a database, where some of it may be toxic and some of it not. With a data set, the output of data mining or a report with combined data elements that are individually a mix of nontoxic and toxic can become toxic in aggregate. Focus on identifying the toxicity of data itself, not data sets, to address this concept of toxic transference. If a piece of data is toxic on its own (e.g., credit card number, shipping address) and this data is included in a data set, it will transfer its toxicity when you combine and use it with other nonsensitive data (e.g., date of purchase order, purchased item).

› **Legacy data classification (i.e., excess data or outdated classifications).** The task of classifying legacy data is a separate project.[9] Establish a process for classifying new data first, and then classifying legacy data will not seem as daunting. But 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 are not handling it as they should. 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 then tackle legacy data classification with brute-force search, eDiscovery solutions, or access control.[10] They can also force the declassification of this data after three years' time or implement a review for reclassification. In most cases, after three years, the value and sensitivity of internal data have diminished, and firms can dispose of it or make it unclassified (public) information.

FIGURE 3 Rethink Data Classification

Traditional solution	Reality	Practical solution
Classify existing information assets.	• It is time-consuming and error-prone. • There is no labeling support.	• Use brute-force search. • Use eDiscovery. • Use access control. • Force declassification after three years.
Educate employees.	Employees don't have time to think.	• Re-engineer workplace so thinking isn't required. • Have a max of three classification levels.
Treat data sets and data the same.	• They are not the same. • Data toxicity is transferable.	Focus on identifying and classifying data, not data sets.
Apply classification labels to new unstructured content.	There is no universal technology support.	• Use dynamic classification from content. • Use access control and ERM for radioactive (highest-level) information. • Use DLP, access control, and best efforts for everything else.

Identify Data Classification Roles

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 4):

› **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 the public domain or 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 on classifying data. Make data classification easy so that data creators can focus on doing their jobs.

- **Data owners.** Data owners may be line-of-business managers, division heads, or the equivalent. If the data resides and is primarily in use within their group, they own it. The data owners review the data creators' tagging of the data. If there is disagreement on labeling, they can start a discussion as to why, and reclassify as needed. 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 owners to provide an additional layer of context for classification, which many automated tools lack today.
- **Data users.** Anyone who has access to this data is a data user. Those who have authorization to handle and use the data are in the best position to provide feedback or answer questions about the data classification tags for data owners and data auditors. Is the classification appropriate and based on how the data is used? Are there circumstances or situations where the data could be handled differently from 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 an equivalent role. The data auditor reviews the data owner's assessment of classification and determines 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 champions.** A data champion is an individual who is 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 (e.g., the chief information security officer). A chief data officer who is responsible for data strategy, including quality, governance, and monetization, may also take on this duty.[11] A data steward who is responsible for data governance, practices, and requirements may be tapped to fill this role too.[12] 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 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's feedback to assess alignment between actual or desired data use	• Is responsible for data strategy • Oversees and provides support for data classification efforts

Approach Data Classification With Two Methods

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 5). Commit to starting data classification for new data first for immediate gains. There are two directions or methods to arrive at a classification label (see Figure 6).

› **Method 1: Classify by existing protections and data controls.** This method of classification involves identifying the data types that exist in the organization and the protections that currently exist for those types to determine their classification level (see Figure 7). Classifying data by how it's protected encourages a better understanding of existing data and how it should be classified for use. This is especially helpful in legacy data classification.

› **Method 2: Classify by data toxicity to determine appropriate data controls.** This method of classification starts at the point of data creation. Classifying data by the data creator's or user's assessment encourages a deeper understanding of how the data is used and valued, providing guidance for its protection.[13] An example workflow for this approach aligns with the three classification levels (see Figure 8). This is helpful for new data classification.

FIGURE 5 There Are Two Types Of Data Classification Projects

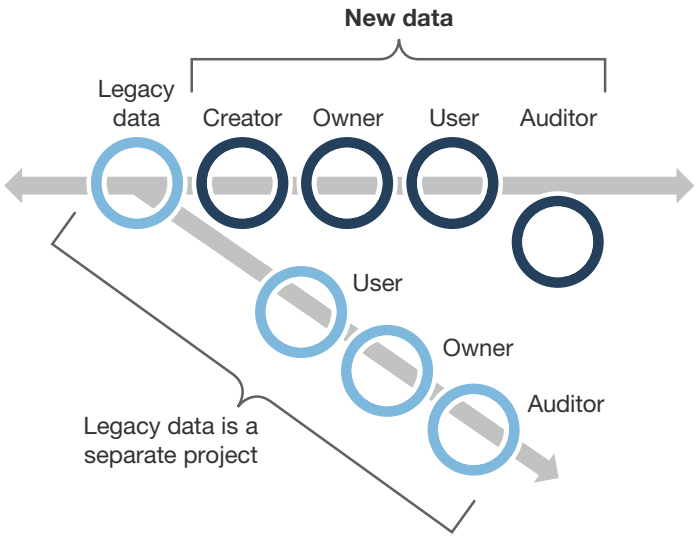


FIGURE 6 Two Methods To Approach Data Classification

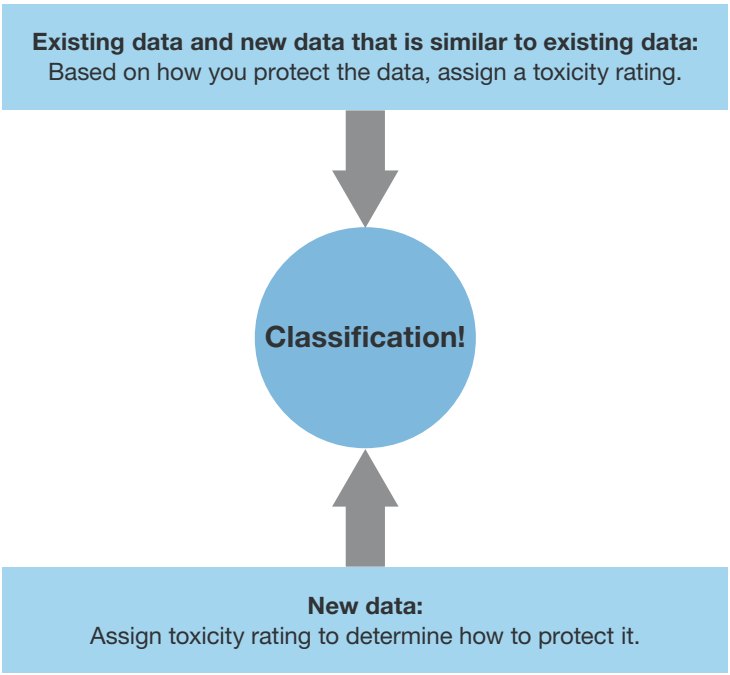


FIGURE 7 An Example Of Classifying Data By How It Is Protected

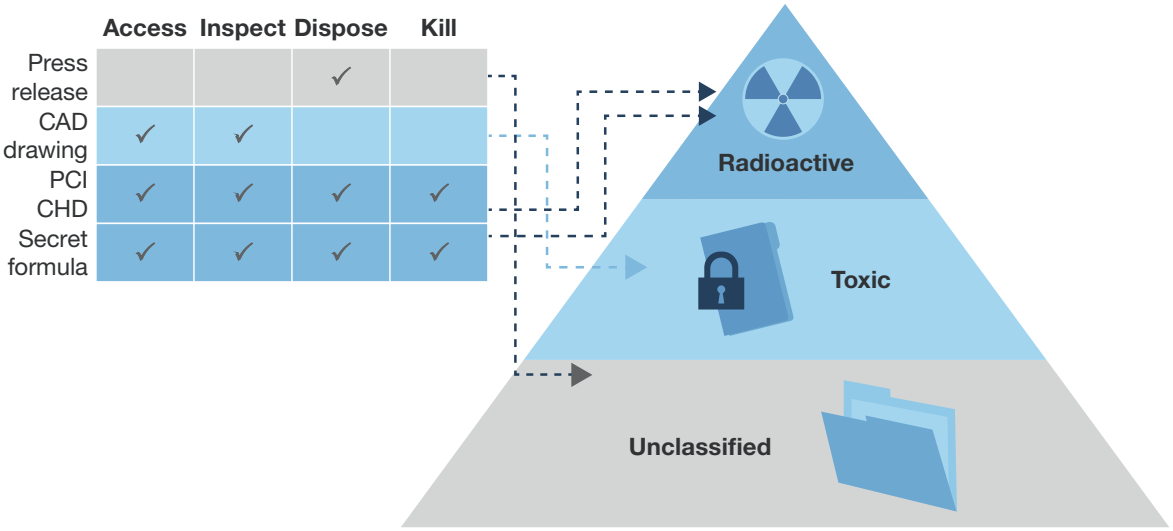
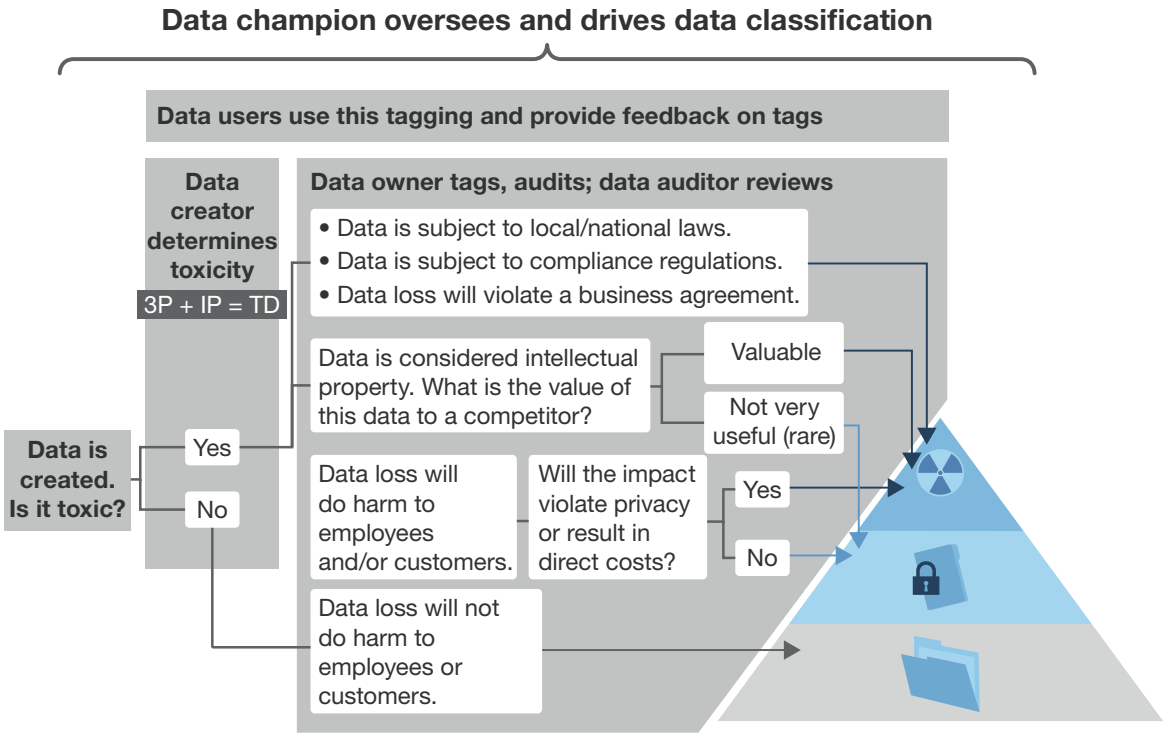


FIGURE 8 An Example Workflow For Classifying Data From The Point Of Creation

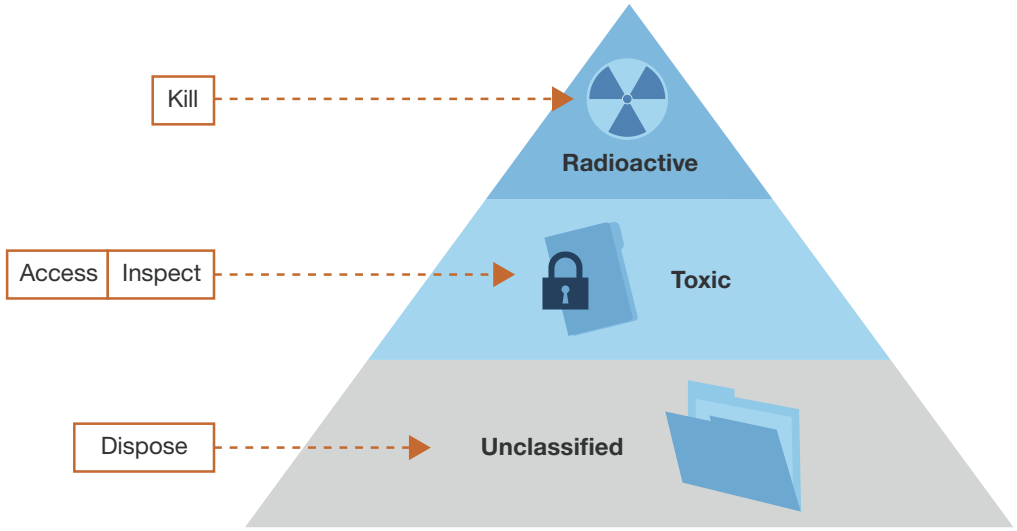


Simplify Classification Schemes

A typical enterprise data classification scheme has anywhere from three to six levels, with a large majority in the three- to four-level range. At an enterprisewide level, simplify classification into three tiers — radioactive, toxic, and unclassified — before considering any sublevels for specific business units if necessary (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. For firms that expect to use privacy as a competitive differentiator, they should treat customer data (e.g., PCI, PHI, and PII) as radioactive. 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, 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 methods for protecting the data, the data is 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 unclassified or public.

FIGURE 9 Simplify Data Classification With Three Levels



Build The Conditions For Dynamic Classification

Treat data as living, not static. Its value is highest at the point of creation and may diminish over time. 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. As this market continues to evolve, look for solutions that can help pull in context about data use and analyze data sensitivity to enable dynamic classification. This may involve contextual analysis, user behavior analysis, machine learning, and other techniques. Two feedback loops help make classification a dynamic rather than static process for your data (see Figure 11):

- **Create a manual feedback loop to provide context and identification information.** At the time of creation, the data creator tags the data as toxic or nontoxic and sends it off to the data environment (i.e., the data's storage location for use). Data users then proceed to access 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 that tools aid and provides much-needed context and identification information for data.
- **Create an automated feedback loop to audit policy enforcement.** This generates a 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 the system. For example, if

certain types of data from a particular business group keep getting flagged as DLP violations but have legitimate exceptions or if the data restrictions are only valid up until a certain point in time, use this knowledge to update classification labels, data controls, or DLP policies.

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

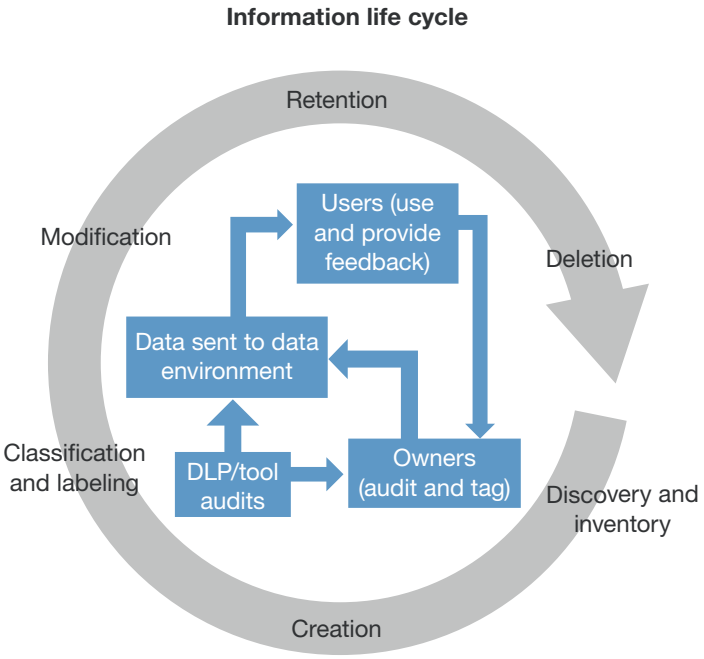
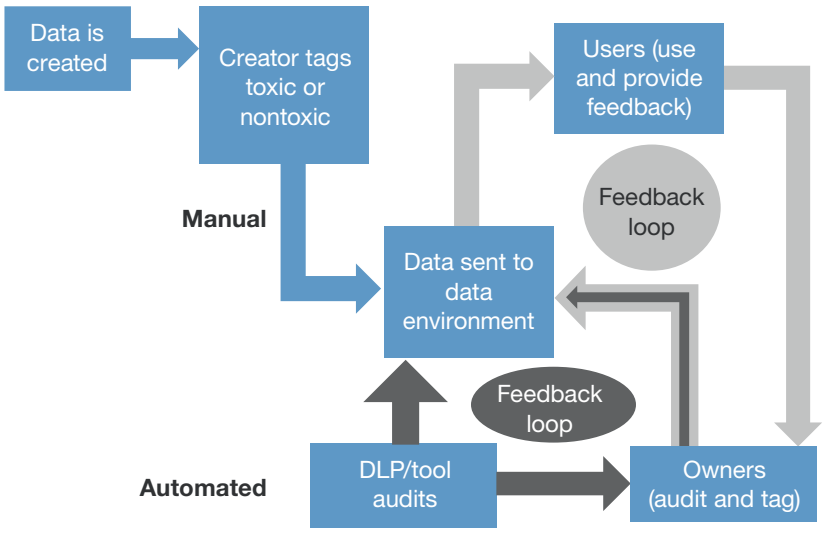


FIGURE 11 Dynamic Data Classification Involves Two Feedback Loops



Recommendations

Now That You've Defined Your Data, Apply That Knowledge Elsewhere

Data discovery and dynamic data classification is not the end, but a beginning, for data defense and a sound approach to privacy. S&R pros: As part of ensuring that ongoing data defense and appropriate handling of data align with your firm's privacy practices, we recommend that you:

- **Consolidate your data.** After engaging in data discovery and classification — defining the data — S&R pros usually find that data exists in unlikely or unexpected places. This is problematic especially if your firm is subject to privacy laws with data residency requirements. This is a great opportunity to consolidate the data, aggregating it so that it meets such requirements or resides in fewer places to limit breach exposure and potentially reduce the scope of other compliance for mandates like PCI. Data consolidation is also a critical process stage in DLP maturity.[14]
- **Create or revisit policies for data collection, use, and protection.** Knowing and understanding the data enables actionable and enforceable data security and privacy policies. The next step is to understand the implications for data handling from storage to appropriate use and disposal and to 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.[15] Specific privacy policies and considerations to address include data collection, purpose, and use.

› **Determine when you should expire data — and identify processes for data deletion.** Data is a living thing. Changes will occur within the business, and the classification level of data will need to align accordingly. For example, data that relates to a merger is extremely sensitive within the time frame leading up to the event but mostly public afterward. Determine how and when to expire data in line with approved corporate retention policies that legal, compliance, or information management peers define. Similarly, to align with privacy policies and regulatory requirements, identify your processes for data deletion. If consumers request that you delete their data, how can they make this request, and how will you ensure that it is done?

Supplemental Material

Survey Methodology

Forrester conducted an online survey fielded in April through June 2015 of 3,543 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.

Forrester's Business Technographics® provides demand-side insight into the priorities, investments, and customer journeys of business and technology decision-makers and the workforce across the globe. Forrester collects data insights from qualified respondents in 10 countries spanning the Americas, Europe, and Asia. Business Technographics uses only superior data sources and advanced data-cleaning techniques to ensure the highest data quality.

Endnotes

[1] S&R pros must control and protect the extreme volumes of data that an 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 S&R pros 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 "The Future Of Data Security And Privacy: Growth And Competitive Differentiation" Forrester report.

[2] 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 tech management, legal, and records management professionals to consider. See the "Market Overview: eDiscovery, Q4 2013" Forrester report.

[3] As data volumes explode, it's becoming a Herculean task to protect sensitive data from cybercriminals and malicious actors while preventing privacy infringements and abuses — intentional and unintentional. Every day, vendors introduce a new product or service that claims to be the cure-all for data security challenges. See the "TechRadar™: Data Security, Q1 2016" Forrester report.

[4] 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 "Know Your Data To Create Actionable Policy" Forrester report.

[5] Complexity also increases as other dimensions such as likelihood and impact of a data breach are taken into consideration.

[6] For example, consider the EU General Data Protection Regulation and EU-US Privacy Shield implications. In addition, the privacy landscape is rapidly changing in Asia Pacific. Globally, privacy laws continue to evolve. See the "Quick Take: EU Gives The General Data Protection Regulation Some Sharp Teeth" Forrester report, see the "Quick Take: Goodbye Safe Harbor, Hello EU-US Privacy Shield" Forrester report, see the "Privacy, Data Protection, And Cross-Border Data Transfer Trends In Asia Pacific" Forrester report, and see the "Forrester's 2015 Data Privacy Heat Map" Forrester 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 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 "Identify And Influence Data Security And Privacy Stakeholders" Forrester report and see the "Job Description: Chief Privacy Officer" Forrester report.

[8] Identify opportunities to introduce a data classification effort alongside an existing business technology initiative. Several companies that Forrester interviewed for this research pointed to new initiatives like a migration to Sharepoint, the introduction of Office365, or a broader information governance effort as additional catalysts for driving data classification.

[9] 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.

[10] 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. See the "Market Overview: eDiscovery, Q4 2013" Forrester report.

[11] The chief data officer role is still a relatively young one across many organizations, and much like the emerging role of CIO was 20 years ago, the core responsibilities of this position are not yet fully established. The new position arose from a growing awareness of the value of data and recognition of an inability to take advantage of the opportunities that it provides — whether due to technology, business, or basic cultural barriers. This new role, however, varies significantly across organizations in terms of where it sits and what it does. See the "Know Your Data To Create Actionable Policy" Forrester report and see the "Top Performers Appoint Chief Data Officers" Forrester report.

[12] In anticipation of the increasing adoption of personal identity and data management 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 that 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 "Building Data Stewardship Is A New Customer Insights Imperative" Forrester report.

[13] 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 data kills (devaluing data through encryption, tokenization, or data-masking technologies). These four levers are the four components of the defend phase of the Data Security And Control Framework. See the "The Future Of Data Security And Privacy: Growth And Competitive Differentiation" Forrester report.

[14] 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, and enforce. See the "Rethinking DLP: Introducing The Forrester DLP Maturity Grid" Forrester report.

[15] 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 "Know Your Data To Create Actionable Policy" Forrester report.

FORRESTER®
CHALLENGE THINKING. LEAD CHANGE.

We work with business and technology leaders to develop customer-obsessed strategies that drive growth.

PRODUCTS AND SERVICES

- Core research and tools
- Data and analytics
- Peer collaboration
- Analyst engagement
- Consulting
- Events

Forrester's research and insights are tailored to your role and critical business initiatives.

ROLES WE SERVE

Marketing & Strategy Professionals

CMO
B2B Marketing
B2C Marketing
Customer Experience
Customer Insights
eBusiness & Channel Strategy

Technology Management Professionals

CIO
Application Development & Delivery
Enterprise Architecture
Infrastructure & Operations
- Security & Risk

Sourcing & Vendor Management

Technology Industry Professionals

Analyst Relations

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.