

# A Blueprint for Data-Centric Security

Applying the Right Protection Directly  
to Data for Increased Security



# Contents

- ▼ What is Data-Centric Security?
- ▼ Data-Centric Security in Action
- ▼ Key Principles of Data-Centric Security
  - ▼ Policies, Rules, and Workflows
    - ▼ Before You Build
  - ▼ Choosing Your Technology
  - ▼ Building and Implementing

Network and device protection continue to consume IT security resources, but they are becoming less and less effective at protecting organizations from cyber threats. Data breaches are now a daily occurrence, creating massive financial burdens for corporations, government agencies, and consumers.

Perimeter breaches will become even more common in the years to come as data volumes continue to grow and IT architecture evolves away from traditional network-based models. In our new digital environment, data-centric security is the only viable option for keeping sensitive information safe.

This paper gives an overview of the key principles of data-centric security, along with best practices for designing and implementing an effective data-centric security solution.

## What is Data-Centric Security?

Data-centric security is a fundamentally different approach for protecting sensitive data from theft or misuse.

Most security technology focuses on where data is, for example, protecting all the data stored on a specific laptop or server, or all the data that crosses a specific network. The problem with this approach is that as soon as data moves somewhere else, another solution is required, or data is left unprotected.

Data-centric security, on the other hand, focuses on what needs to be protected—the files containing sensitive information—and applying the appropriate form of protection no matter where the data happens to be.



“Networks no longer have distinct borders, and perimeter security has failed to keep pace with the advances in technology.”

Nucleus Research

## Data-Centric Security in Action

The defining characteristic of data-centric security is that protection is applied to data itself, independent of the data's location. To be effective, this must happen automatically; sensitive information should be identified as soon as it enters an organization's IT ecosystem, and should be secured with policy-based protection that lasts throughout the data lifecycle.

A typical implementation of data-centric security consists of software agents installed on every IT asset where sensitive data might be created or stored—laptops, desktops, servers, mainframes, mobile devices, and elsewhere. These agents are controlled by a centralized management console, where administrators define the appropriate form of protection for each data type and use case.

Each time a file is created or modified, the system scans the file to determine whether it contains sensitive information, and automatically applies the appropriate protection. End users may be given the ability to modify these actions manually, but are otherwise not involved in the process. Protected data remains available to authorized users, but cannot be accessed by unauthorized users, even when files travel outside the company network.

When implemented on an organization-wide scale, data-centric security reduces or eliminates the impact when network and device protections inevitably fail, while at the same time removing data silos and other internal obstacles.



## Key Principles of Data-Centric Security

Each organization requires a unique solution—one tailored to fit the company's threat exposure and business needs. However, all successful implementations of data-centric security have certain characteristics in common: They are tightly controlled from a centralized management system, they provide coverage across the entire organization without security gaps, they rely on automation rather than manual intervention, and they are adaptable enough to grow and change along with the organization.

Every data-centric security solution must be designed with these key principles in mind, or the finished product will fall short of its goals.

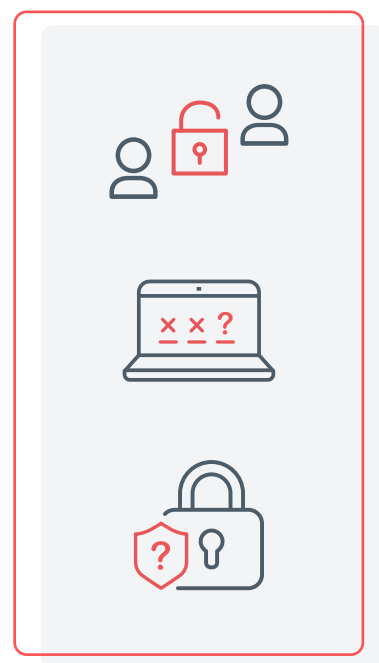
## Centralized Control

Centralized management is essential to ensure that data is protected according to the organization's security policies, and that data remains available for appropriate use.

Most organizations, even those that have not adopted the data-centric model, already have some data-level protection in place. Typically, this takes the form of user-applied file encryption—when employees encrypt files with passwords before sharing them with colleagues or external recipients.

Though very common, this approach creates three significant problems:

- Employees might neglect to apply protection when it is needed, or might choose methods that don't provide adequate protection.
- When they do encrypt their files, employees must then find a way to share the passwords (the encryption keys) with the recipients, which typically requires the use of unencrypted email or another unsecure method.
- User-applied encryption leaves employees, rather than administrators, in control of encryption keys. When keys are not available (for example, when employees fail to share them, forget them, or leave the company without providing them), the organization can permanently lose access to critical data.



When properly implemented, data-centric security gives the organization complete control over its sensitive data from the moment that each file or database record is created. Access to protected data can be granted or revoked at any time, and all activity is logged for auditing and reporting.

## Automation

Automated workflows are the key to success in data-centric security. End users do have a part to play in protecting an organization's information, but they should not be expected to shoulder the burden of evaluating and securing the large and constantly growing volumes of data they handle each day.

Automation takes user error out of the equation and allows employees to do their jobs without interruption and without jeopardizing the company's data.

An organization's security technology must apply its data protection policies in real time, across the entire enterprise, without user intervention. This requires technology that continuously monitors file activity, and automatically applies the appropriate protection as soon as sensitive data appears.

## Policies, Rules, and Workflows

As in any rapidly evolving industry, cybersecurity's vocabulary is always changing. However, the following terms are commonly used to describe the distinctions between an organization's high-level security requirements and the ways those requirements can be implemented in the "real world" of the organization's IT ecosystem.

**Policies** are written documents that define an organization's standards and goals for data security.

Without automated technology, an organization must rely on employees to remember and follow its data security policies.

**Rules** are specific software configurations that apply an organization's policies to different use cases.

For example, if an organization's written policy states that all files containing credit card numbers must be encrypted, it can create rules in its software to specify which devices and locations need to be monitored for files containing card numbers, and which encryption methods and keys should be used to protect them.

**Workflows** are the internal processes used by software to carry out the actions specified by rules.

Workflows describe each step taken by a system in a given use case, including monitoring file activity, scanning for sensitive data, applying classification tags, and applying encryption or other forms of protection.

they don't want any surprises.

## Adaptability

Data-centric security is not a “one size fits all” proposition. Within a single organization, there can be dozens of security policies, hundreds of data types, and thousands of use cases. Some data might require encryption, while other data may need to be masked, deleted, quarantined, or left as is.

An effective security strategy will be tailored to meet the organization’s unique requirements, while accommodating changes in those requirements over time. Organizations must have confidence that they can add and remove infrastructure, change business processes, and create new partnerships, without having to rebuild their data security solution each time.



## Before You Build

It’s understandable that many organizations struggle to get the resources or budgets needed to achieve full compliance every year. But failing to properly scope (or pass) an assessment costs even more time and money, so getting it right the first time is the goal. However, too many companies only think about PCI compliance once or twice a year when it’s time for an assessment—and therein lies the problem.

According to Verizon’s 2020 Payment Security Report, an astounding 72 percent of all organizations assessed on PCI DSS compliance failed an interim validation.<sup>6</sup> That means that, although they had achieved 100 percent compliance in a previous assessment, only 28 percent were able to maintain that level of compliance in between assessments. Perhaps more worrisome: In the past decade’s worth of these annual reports, the best performance was reported in 2017, when 55 percent of organizations were found in full compliance.<sup>7</sup> That number has been declining every year since.

## Conduct a Data Risk Assessment

Risk assessments are a fundamental step in any cybersecurity effort and are called for in a variety of guidelines and mandates, including the National Institute of Standards and Technology’s cybersecurity framework, and New York’s cybersecurity law for financial services companies.

Even if your organization has conducted cyber risk assessments in the past, it is best to make a fresh assessment before beginning a data-centric security implementation. This will ensure that you account for any recent changes to IT infrastructure or business processes and identify any data-specific issues that may not have been documented in previous assessments.

Specifically, your data risk assessment should focus on the following elements:

- The types of data being created and acquired by users, applications, and automated processes
- The use cases for each type of data and each user group
- The risks associated with each data type and use case
- The role and effectiveness of any existing data protection technology
- Any applicable government or industry mandates relating to data security, such as GDPR, PCI-DSS, HIPAA, or NYCRR 500, and the consequences of noncompliance.



With this information in hand, you will be prepared to set your organization's data security priorities and define policies that will guide you as you choose your new technology and build your data-centric security solution.

## Define Your Data Security Policies

Written security policies are the foundation for any data-centric security strategy. In addition to communicating the organization's security standards to administrators and end users, written policies provide assurance to regulators, corporate boards, and customers that the company understands the importance of information security.

Most organizations' written policies focus on protocols for device access and network access, because those are the areas that traditional security technology can address. With a data-centric approach, however, companies gain the ability to enforce specific standards for how each type of data should be managed and protected, regardless of where the data resides.

If your organization's written policies do not already provide guidance on how different forms of data should be treated, the policies should be expanded to answer the following questions:

- What are the organization's data security compliance obligations?
- What are the expectations of the organization's customers, employees, board members, auditors, and government regulators?
- Which types of data are considered sensitive?
- Which user groups or profiles should have access to each type of sensitive data?
- What forms of protection or remediation are required for different data types?
- What are the required timeframes for data retention or deletion?

## Choosing Your Technology

Once your organization has documented its data security goals and requirements, the next step is to choose a technology that can help it meet those objectives.

To make the right decision, your organization will need to consider its capacity for implementing new technologies, the basic architecture of a data-centric security solution, and the capabilities it will need in order to enforce your written policies across the enterprise.

### Platforms vs. Point Solutions

Point solutions—products that address only one or a few use cases—are commonly used for network and device security, but are not well suited to the data-centric security model. Using multiple independent products to apply data-level protection can create data silos and security gaps, defeating the purpose of the implementation.

Rather than purchasing multiple products from multiple vendors, organizations are better served by implementing a comprehensive data security platform—a solution that integrates the full set of capabilities they require and provides options for adding or creating new capabilities in the future.

Not only does the platform approach simplify administration by giving managers a single point of control for all data security activity, it simplifies the process of expanding or enhancing the solution in the future to address changes in its infrastructure or business processes.

## Basic Elements

A data-centric security solution typically consists of a management console and an array of software agents, along with any supporting infrastructure the organization might require.

- The management console is the central component of the solution, where administrators convert the organization's written policies into automated rules and workflows that will be carried out by agents across the enterprise. The management console should also provide reporting and auditing tools that allow the organization to monitor activity and identify emerging threats.
- Agents installed on laptops, desktops, servers, and other IT assets are responsible for monitoring file activity and taking action on sensitive data. Agents must remain in regular communication with the management console in order to receive policy updates and to deliver data for logging and auditing.
- Supporting infrastructure might consist of a hardware appliance to host the management console, as well as any additional elements such as hardware security modules, random number generators, or KMIP (Key Management Interoperability Protocol) connectors.



## Basic Elements

In order to address the variety of data types and use cases required by a large organization, a security solution must provide a wide range of capabilities. Each of the features and functions listed below is likely to be necessary for a large corporation or government agency.

- Policy management is the primary function of an administration console. An organization's written security policies need to be translated into system settings, automated rules, and other configurations that control how sensitive data is identified and protected. The more complex an organization's use cases, the more important it is that the administration console provides an intuitive way to create and organize those settings.

- Data discovery involves monitoring file activity and automatically scanning new or modified files to determine if they contain sensitive data. As each organization will have its own definition of “sensitive data,” the solution must allow administrators to define criteria that will identify a file as sensitive.
- Classification is the process of tagging files with metadata and visual labels to indicate their contents and appropriate use. In most cases, classification should be automatically applied based on the results of a data discovery scan. Some organizations also give end users the ability to add, remove, or change file classifications.
- Persistent encryption is the strongest form of data protection, rendering data inaccessible to unauthorized users while allowing authorized users to decrypt and read a file’s contents. A robust key management feature is essential to ensure that encrypted files are accessible to only the proper users at all times.
- Data masking and redaction involve replacing sensitive data with non-sensitive data (for example, replacing credit card numbers with strings of asterisks), or removing sensitive data from a file. These approaches are often used when files contain sensitive information along with other information that needs to be made widely available.
- Moving and quarantining data is often necessary when files containing sensitive information are saved in inappropriate locations. For example, if hospital employees are not permitted to save patient health information on their laptops, a data-centric security solution can be configured to copy patient files to a secured server and delete the files from the laptops where they appeared.
- File deletion may be necessary in cases in which a file contains data that should not be saved or stored in a specific location, or should not exist within the organization at all.
- Auditing and reporting are essential for detecting threats and demonstrating compliance with internal and external mandates. A solution’s management console should provide detailed reporting on which files contain sensitive data, where the files are located, and who has been granted access to them.
- Integrations with other key elements of the organization’s IT infrastructure help streamline workflows and ensure that all sensitive data is managed appropriately. A data security solution should be able to work with the organization’s ERP systems, proprietary applications, productivity tools, and other technology.

# Building and Implementing

Many organizations take a phased approach when implementing data-centric security, addressing their most critical security risks (and/or their most straightforward use cases) first, before expanding the solution to additional user groups, operating systems, and business units.

Typically, the implementation process consists of defining criteria that will be used to identify sensitive data, creating automated rules and workflows that apply the organization's written policies to different data types, and deploying agents to enforce the organization's policies wherever sensitive data can be created, acquired, modified, or stored.

## Defining Sensitive Data

Most organizations categorize information as “sensitive” if its ownership or use is restricted by a government or industry mandate, or if it cannot be made public without damage to the organization's reputation and ability to compete.

An organization's general definition of sensitive data should be established in its written policies, but security administrators typically need to create more detailed definitions in order to correctly identify files that require protection.

Criteria for identifying data can be defined in a number of ways:

- Patterns are sequences of characters that match specific formats, such as Social Security numbers or credit card numbers. To prevent false positives, discovery scanning technology needs to incorporate the specific algorithms that are used to generate these data types.
- Dictionaries are lists of specific terms (such as “patient” or “confidential”) that may be found in files containing sensitive data.
- Regular expressions are logical statements that can be used to identify certain types of data and are particularly useful for detecting data that one organization might consider sensitive, but is not generally considered sensitive by other companies.

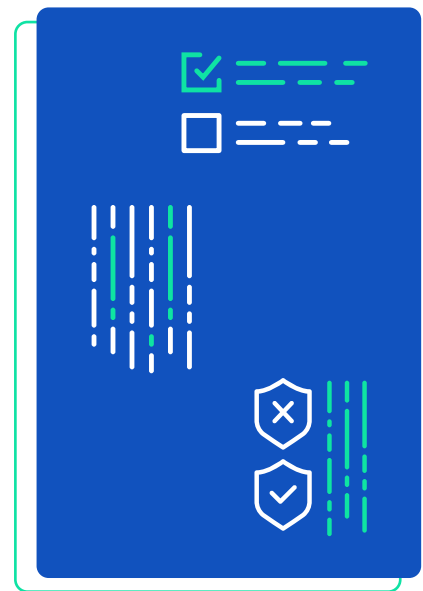


## Creating Rules

Rules are the mechanisms for applying the organization's written policies to its sensitive data, and define what should happen to files containing sensitive data in different situations.

Data protection rules can take many different forms, depending on the organization's IT infrastructure and approach to granting access to sensitive data. In general, though, most rules follow a general pattern:

1. Define the data type to which the rule applies (for example, credit card numbers)
2. Define the conditions under which the rule should be executed (for example, when a file containing credit card number is saved on an employee laptop)
3. Define the action to be taken when the rule conditions are met (for example, encrypt the file with a specific key, or copy the file to a quarantined location and delete the original)



As with the definition of sensitive data, an organization's list of data protection rules will likely be more detailed than the general standards described in its written policies. A large organization may need to create hundreds of rules to address its full range of data types and use cases.

After an organization's security administrators have defined the data types requiring protection and the rules to be followed when sensitive data is detected, the remaining step is to deploy software agents that will enforce those rules across the enterprise.

## A New Definition of Security

Once an effective data-centric security solution is in place, it becomes impossible for sensitive data to exist in violation of the organization's security policies.

Files are automatically scanned upon creation or modification, and are given appropriate protection without the need for action by end users. If a protected file is shared in an unsecure cloud location, mistakenly emailed to an unauthorized user, or stolen by an intruder, the data contained in the file remains safe from exploitation.

Data-centric security is the only approach that provides this level of security, and the only approach that can deliver meaningful protection against today's constantly evolving cyber threats.

PKWARE's automated data security platform allows organizations to detect, classify, and protect sensitive data on every enterprise operating system.

With PKWARE, your organization can meet its compliance obligations and keep data safe from internal and external threats, without disrupting the way you do business today.

---

## About PKWARE

PKWARE offers the only data discovery and protection solution that locates and secures sensitive data to minimize organizational risks and costs, regardless of device or environment. Our ultra-efficient, scalable software is simple to use on a broad range of data types and repositories, enabling precise, automated visibility and control of personal data, even in the fastest-moving, most complex IT environments. With more than 1,200 customers, including many of the world's largest financial institutions, retailers, healthcare organizations, and government agencies, PKWARE continues to innovate as an award-winning global leader in data discovery, security, and compliance. To learn more, visit [PKWARE.com](https://pkware.com).

## **Enterprise Wide Policy Management**

The PKWARE Enterprise Manager provide a single point of control for data protection activity across the entire organization

## **Simple Workflow**

With PKWARE, data protection is automated for end users and easy for administrators to manage

## **Easy Implementation**

PKWARE supports a variety of deployment options, enabling organizations to implement their data protection solution without time-consuming changes to infrastructure and workflows

## **Protection Without Gaps**

PKWARE works on every enterprise operations system and provides persistent protection that remains with data even if it's copied or shared outside organizations

## **Integrated Discovery, Classification and Protection**

No other solution has the capability to find, classify and protect data in a single automated workflow

## **Multiple Protection and Remediation Options**

Organizations can take a policy-based approach to data protection and choose from action including persistent encryption, quarantine, masking, and deletion.



**PKWARE.com**

866-583-1795

201 E. Pittsburgh Ave.  
Suite 400  
Milwaukee, WI 53204

NASSCOM



Microsoft  
Partner

