

PK Discovery

Maintaining Continuous Visibility of
Personal Data Across the Enterprise



Contents

- ✔ What is PK Discovery?
- ✔ Why Invest in PK Discovery?
- ✔ Why Is PK Discovery Needed Now More Than Ever?
- ✔ What Makes PK Discovery the Best and Safest Choice?
- ✔ What Platforms and Data Repositories Can PK Discovery Scan?
- ✔ Time-Tested and Customer-Validated, PK Discovery Surpasses Them All

Introduction to PK Discovery

PK Discovery is the leading solution in providing personal data discovery for protection to minimize organization risks and costs, regardless of device or environment. PK Discovery is a fully automated software solution for discovering all personal, confidential, and otherwise sensitive data any company has amassed. PK Discovery locates, identifies, indexes, and reports the full inventory of a company's personal data, from the broadest range of data repositories and servers—whether on premise or in the cloud—to end point folders and files. It is straightforward to deploy and use—no coding required—and delivers unmatched power, precision, and scalability. It uses a sophisticated machine learning approach to identify personal data of any type with unbeatable breadth and depth of coverage and support. In use for nearly 15 years at Fortune 500 companies across the world, it's the most proven and technologically advanced solution the market offers.

Why Invest in PK Discovery?

Data volumes are growing exponentially. New data types, formats, and repositories are being created all the time. Data is being shared in the cloud and around the globe in millions of digital channels. Every electronic device—at home, at work, on the go—is constantly transmitting personal information, and companies need to locate and properly label and treat all of it.

Data management, information governance, and regulatory compliance are core drivers that ultimately move companies toward adopting and investing in personal data discovery technology. Enterprises in highly regulated industries already recognize personal data discovery as an absolute necessity. Due to PK Discovery's precision in personal data discovery, enterprise after enterprise selects our technology in



Use data discovery and classification to gain greater visibility and understanding of your organization's sensitive data, secure sensitive data with appropriate controls and policies, and support compliance, privacy, and ethical data use.

—Forrester Research,
October 19, 2020

competitive proofs of concept. Every company that tries PK Discovery rapidly and readily finds more hidden sensitive data than with any other solution, and in places they'd never expected to.



Functionality such as the breadth of data-type coverage, data identification scope, data identification techniques, and tagging and labeling capabilities—on top of our ease of use—sway those companies to choose PK Discovery. PK Discovery rapidly and thoroughly sets the foundation for consistent, policy-driven governance, data management, and compliance.

Another strong attraction for companies choosing to invest in PK Discovery is that our company is partner-led, particularly in large enterprise arenas. Enterprise customers tend to require customization in their IT solutions and want to work with consulting and technology service providers already do business with.

Our partners choose to work with PKWARE due to our wide range of platforms and ability to discover personal data in structured, semi-structured, and unstructured data sources. We make customizations far smoother and simpler for our partners who need to provide customer accounts reliability. Consulting services as well as technology integrations smooth onboarding and operations. On all counts, PK Discovery empowers business rather than hindering it.

Why Is PK Discovery Needed Now More Than Ever?

PK Discovery is the only solution delivered with over 100 out-of-the-box policies ready for companies to use day one. A company can buy PK Discovery, set it up quickly, and be discovering unknown personal data within minutes. That efficiency translates into real money. All of the verticals noted above will continue to be more and more regulated by emerging laws across the country and the globe. Privacy laws have already heavily regulated healthcare and life sciences; now, given the massive activity in research and treatment of illnesses due to coronaviruses, their use of personal data discovery is anticipated to explode as time goes on.

Notably, in 2021 and as we advance, the rise in companies' remote workforces resulting from lockdowns and improved productivity, and the ongoing artificial intelligence (AI) and machine learning (ML) accelerated supply and demand of access to real-time data all factor heavily into enterprise investments of personal data discovery going forward. According to MarketsandMarkets™, the global personal data discovery market size is expected to grow from USD \$5.1 billion in 2020 to USD \$12.4 billion by 2026, at a Compound Annual Growth Rate (CAGR) of 16.1 percent.

There will be both emerging and evolving governmental agencies and regulations impacting companies as we look forward. For example, on November 4, 2020, California voters passed the California Privacy Rights and Enforcement Act (CPRA) on the heels of the California Consumer Privacy Act (CCPA). Data that companies acquire in 2021 and continue to use or share into 2022 will be looked at under this law's purview. The CPRA grants additional consumer rights, including:

- Right to limit use, sharing, and disclosure of sensitive personal information
- Right to correction of personal information
- Right to access information regarding the company's automated decision making
- Right to opt-out of automated decision making
- Technology right to restrict sensitive and personal information audit obligations

The law includes specific provisions for the use of children's personal data and employees' personal data. The passing of this law also established the California Privacy Protection Agency that will be partially funded by collecting fines for infractions on the legal use of personal data.

It becomes necessary and establishes a competitive advantage to rapidly and thoroughly discover all the sensitive and personal data a company has with the best technology available, know how to index and treat the data, protect it appropriately, and remain reliable and viable in their markets.



What Makes PK Discovery the Best Choice?

Simple to Deploy and Use for Accelerated Time to Value

Unlike other solutions requiring extensive integration or professional services work to get started, PK Discovery does not require a single line of code to install and deploy. It is a policy-based solution capable of discovering any data in multiple languages that is defined as sensitive, personal, or confidential by the enterprise. It requires minimal

to no effort to get up and running, and within minutes it can deliver fine-grained and aggregated optics into exactly where personal data resides in enterprise data stores, endpoints, and fileshares. Users can take advantage of more than 100 out-of-the-box policy templates for PII, PCI, HIPAA, GDPR, CCPA, CPRA, LGPD, PCI DSS, and other regulatory elements, or define custom policies with unlimited types and numbers with just a few clicks.



Intelligent Discovery and Detection for Precise and Thorough Results

Unlike other solutions that use metadata analysis or probabilistic heuristics to infer or guess where it might locate personal data, PK Discovery scans the data itself at the smallest element level. It performs deep content inspection of structured, semi-structured, and unstructured data using special techniques that incorporate dictionary-based and weighted keyword matches, patent-pending neural-like network (NLN) technology, intelligent contextual analysis, and advanced machine learning to discover personal elements more accurately.

PK Discovery can find and report the exact quantity and location of personal data by first finding all the candidate elements that may be personal, such as a 16-digit number. This particular candidate may or may not be a credit card number, but certain validation functions, both mathematical and logical, help determine that. Then PK Discovery uses proprietary algorithms to determine and classify exactly which candidates are of what personal type. Certain data elements are more complex, such as addresses and names, which require additional positional and ordering checks. In these cases, PK Discovery uses a grammar-based approach to classify a group of candidate matches.





Finally, within vast data sets, certain elements are ambiguous. Although PK Discovery produces no false negatives, it must first remove all ambiguities to make an accurate determination of the personal data type and ensure zero false positives. So, PK Discovery performs contextual analysis, taking advantage of machine learning to get smarter. For example, a Card Verification Value (CVV) of a Visa credit card account is three-digit number; if all three-digit numbers were classified

as CVVs, scans would result in many false positives. PK Discovery scans and looks for credit card numbers, specific keywords, and other qualifying information that pertain to or are in the vicinity of three-digit numbers before correctly making or not making any CVV classification.

PK Discovery combines of all these techniques with highly filtered, accurate reference data, inclusion and exclusion lists, and user-defined validation functions to deliver the most unambiguous and thorough discovery of personal data in the industry.

Examples of Personal Data PK Discovery Finds

- | | |
|---------------------------|--------------------------------|
| ■ Full Name | ■ Geographical Indicators |
| ■ Alias | ■ Device Serial Number |
| ■ Last Name | ■ Device Identifier |
| ■ First Name | ■ Medical Record Number |
| ■ Maiden name | ■ International Identification |
| ■ Driver's License Number | ■ Place of Birth |
| ■ License Plate Number | ■ Date of Birth |
| ■ Passport Number | ■ SWIFT Code |
| ■ Employment Information | ■ Financial Account Number |
| ■ Taxpayer ID | ■ Credit Card Number |
| ■ Insurance Number | ■ Financial Information |
| ■ Geographical Indicators | ■ National ID Number |
| ■ Street Address | ■ URL |
| ■ Zip Code | ■ IP address |
| ■ City | ■ Email Address |
| ■ State/Province | ■ Certificate/License Number |
| ■ County | ■ Fax Number |
| ■ Country | |

Powerful, Proven Scalability to Grow with Your Business

The scale at which PK Discovery can scan data repositories for personal data is unparalleled, scanning an entire repository or partial and sample data sets. It is uniquely optimized to leverage multithreading and concurrent processing, whether deployed in a centralized or highly distributed infrastructure.

Some of the world's largest infrastructure deployments have proven PK Discovery with systems holding dozens of petabytes of data. Even in the unstructured world of Hadoop, a typical PK Discovery enterprise scan is often on the order of hundreds of terabytes. The customers may also require running multiple scans adding up to tens of petabytes—and PK Discovery completes the scans in a matter of minutes or hours, not weeks or months.



Enterprise-Grade Quality, Flexibility, and Support

PK Discovery has undergone rigorous penetration testing and security certifications in Fortune 500 companies worldwide. Delivering an unlimited number of scans, PK Discovery is extremely cost-effective and supports manual and fully automated processes using commandline or RESTful APIs. Customers benefit from its high-availability deployment architecture, designed for zero downtime with disaster recovery and 24/7 customer support. PK Discovery licensing costs and models are flexible, meeting operational requirements of enterprises of any size.

Broad Platform Support for the Most Comprehensive Coverage

Deploy PK Discovery in siloed groups and departments within a company or at an enterprise-wide level. We recommend and consider best practice enterprise-wide deployment for personal data management, governance, and compliance. PK Discovery finds and reports personal data in all flavors of Hadoop, relational databases, on-premises file shares, object-based cloud stores, or servers, folders, and files on any device. There is no other solution delivered out-of-the-box that provides the breadth of platform-scanning capability at the element level as PK Discovery.



What Platforms and Data Repositories Can PK Discovery Scan?

Relational Databases (RDBMS) and Structured Data Stores

- Amazon Aurora
- Microsoft SQL Server
- Greenplum
- MySQL
- IBM Db2
- IBM DB2 for z/OS
- IBM Power Systems
- Oracle
- PostgreSQL
- Sybase MariaDB

Data Warehouses

- Amazon Redshift
- Google BigQuery
- IBM Netezza
- Teradata

Big Data Hadoop Platforms

- Amazon Elastic MapReduce (EMR)
- Cloudera
- Google Cloud DataProc
- Hortonworks
- MapR
- Microsoft Azure HDInsights

NoSQL Databases

- Cassandra

Cloud Object Stores

- Amazon Simple Storage Service (S3)
- Google Cloud Storage
- Microsoft Azure Blob Storage
- Microsoft Azure Data Lake Storage

Inflight Data Transfers

- Via API Library

On-Premises File Servers

- Linux
- Microsoft Windows

Endpoints

- Windows
- macOS
- Ubuntu
- CentOS

File Servers

- Windows Server
- Ubuntu
- CentOS

Time-Tested and Customer-Validated, PK Discovery Surpasses Them All

PK Discovery is the culmination of nearly 15 years of subject matter expertise in sensitive data discovery. Throughout the years, our software has proven itself in the most complex and expansive data environments spanning the world, discovering unknown personal data hidden in petabytes of data, on premises, in the cloud, and mixed.

No in-house or marketed solution matches PK Discovery's rapid accuracy and massive coverage. Other data discovery solutions at times require manual intervention or assistance, which drastically slow the operation. A lack of precision leads to underreporting stored personal data and regularly reporting false positives and false negatives—ultimately translating to a false sense of security for customers. Because other solutions cannot match the scale that PK Discovery offers, they often prove insufficient, unreliable, and costly.

Gain the ultimate control and visibility of personal data with PK Discovery. Once PK Discovery discovers and indexes personal data for appropriate and consistent protection:

- Executives have assured compliance and certainty to avoid costly breaches
- Innovators and decision-makers receive access to complete data sets for significant competitive advantage
- Development and analytics teams have the freedom to leverage the power and efficiency of the company's complete data for insights toward accelerated market value

Perform your own proof of concept with PK Discovery today. Find out how market leaders in the key verticals benefit from using PK Discovery to discover, index, and prepare personal data for protection and competitive leverage.

“Evolving regulations are expected to drive the adoption of the sensitive data discovery solutions and services for structured and unstructured data. Companies and organizations need to ensure their sensitive data discovery solutions are not only secure, but also support data analysis for strategic business decisions.

—Forrester Research, October 19, 2020

Enterprise Wide Policy Management

PK Protect provides a 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 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.



PKWARE.com

866-583-1795

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

