DATAGUISE

Dataguise Discovery

THE “ORIGINAL” SENSITIVE DATA DISCOVERY SOLUTION

DATAGUISE

DATAGUISE DISCOVERY

Security and risk pros can't expect to adequately protect customer, employee, and sensitive corporate data and IP if they don't know what data exists, where it resides, how valuable it is to the firm, and who can use it.

— Forrester Research
Rethinking Data Discovery and Classification Strategies
July 10, 2018

The "Original" Sensitive Data Discovery Solution

Dataguise Discovery is a fully automated software solution for discovering and detecting all personal, confidential, and otherwise sensitive data amassed by an enterprise. In use at Fortune-500 companies since 2007, it is, by far, the most technologically advanced and proven solution available today. Dataguise is extremely simple to deploy and use—no coding required—yet delivers unmatched power, precision, and scalability. It uses a sophisticated, data-centric approach to identify sensitive data of any type in the widest range of data repositories, on premises and in the cloud. For unbeatable visibility of sensitive data across the enterprise, with both breadth and depth of coverage, Dataguise Discovery is the clear choice.

WHO NEEDS SENSITIVE DATA 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.

For many reasons—and especially for enterprises in financial services, insurance, healthcare, retail, government, or other highly regulated industries—sensitive data discovery is an absolute necessity. But the best reason is its ability to help organizations make informed decisions about governance, risk, and compliance that ultimately empower the business, not hinder it.

Sensitive data discovery is indispensable to any data steward responsible for protecting the privacy and confidentiality of sensitive personal or corporate information, such as a:

- Data Protection Officer (DPO),
- Chief Information Security Officer (CISO),
- Chief Information Officer (CIO),
- Chief Privacy Officer (CPO),
- Digital Risk Officer (DRO),
- Chief Data Officer (CDO),
- Chief Compliance Officer, or
- Chief Marketing Officer (CMO).

Any one (or, more often, a combination) of these stakeholders needs sensitive data discovery to make smarter business and technological decisions about where, when, why, how, and by whom data is being collected, stored, or used. Without it, there are potentially dire consequences and often-irreversible damages to both brand and bottom line.

WHY IS IT NEEDED NOW MORE THAN EVER?

Data privacy has always been a concern for both individuals and corporations, since the first byte was born. Over the years, various governmental regulations and industry standards have been created to address specific types of sensitive data—such as the Payment Card Industry Data Security Standard (PCI DSS) or the Health Insurance Portability and Accountability Act (HIPAA). In the United States alone there are hundreds of state and federal laws regarding digital commerce or communication that may apply to an enterprise, but there is no single overarching, cohesive law like the European Union's General Data Protection Regulation (GDPR)—at least, not yet.

The GDPR is the most comprehensive data privacy law to date. The California Consumer Privacy Act (CCPA) is similar. Sensitive data discovery is the fundamental first step to compliance with most data privacy regulations, but these laws make it even more necessary with the requirement to give individuals access to all personal data held about them without undue delay. The GDPR also requires some organizations to appoint a data protection officer, for whom enterprise-wide sensitive discovery and detection will prove to be invaluable. In addition, the GDPR mandates data privacy by design and by default, meaning that sensitive data discovery must be an ongoing, integral part of any business process that involves data—especially as it flows among users, between enterprises, across international borders, and into the cloud. Keeping track of it all manually will be impossible, and fines for non-compliance could be substantial.

If regulatory compliance isn't enough to convince you that your enterprise needs a sensitive data discovery tool, just scan the headline news to read about the latest data breaches. Companies are being called out for their mishandling of sensitive data, and they're paying for it, in terms of brand reputation, customer relationships, and profitability.

> *Until now, consumers have been willing to lend their data (or have unknowingly given it away) to get convenience or information in return. Once they fully realize the consequences of this bargain they will be looking to government and business to safeguard data and hand control back to them, the customer.*
>
> — Forbes
>
> *Data Privacy Will Be The Most Important Issue In The Next Decade,*
> November 26, 2019

DATAGUISE

TOP 5 REASONS TO CHOOSE DATAGUISE FOR SENSITIVE DATA DISCOVERY AND DETECTION

1

Simple to Deploy and Use for Accelerated Time to Value

Unlike other solutions that require extensive integration or professional services work just to get started, Dataguise Discovery does not require a single line of code to install and deploy. It is a policy-based solution, capable of discovering any data defined as personal, sensitive, or confidential by the enterprise, in multiple languages. 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 sensitive data resides in enterprise data stores. Users can take advantage of more than 80 out-of-the-box policy templates for PII, PCI, HIPAA, and GDPR elements, or define unlimited types and numbers of custom policies with just a few clicks.

WHAT SENSITIVE DATA CAN DATAGUISE FIND?

(Representative list—contact us for more)

- Full name
- Alias
- Last name
- First name
- Maiden name
- Driver's license number
- License plate number
- Taxpayer ID
- Street address
- Zip code
- State / Province
- County
- City
- International Identification Numbers
- SWIFT code
- Financial account number
- Credit card number
- Financial information (e.g., salary, account balance)
- Passport Number
- Insurance number
- Device serial number
- Device identifier
- Geographical indicators (location, movement)
- Medical record number
- Date of birth
- Place of birth
- Certificate/license number
- Telephone number
- Fax number
- Employment information
- Passport number
- National ID number
- URL
- IP address
- Email address

DATAGUISE

2

Intelligent, Precise Discovery and Detection for More Accurate Results

Unlike other solutions that use metadata analysis or probabilistic heuristics to infer or guess where sensitive data might be located, Dataguise scans the data itself—at the 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 sensitive elements more accurately.

Dataguise Discovery is able to find and report the exact quantity and location of sensitive data by first finding all the candidate elements that may be sensitive, 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 Dataguise uses proprietary algorithms to make a deterministic classification of exactly which candidates are of what sensitive type.

There are certain data elements that are more complex, such as addresses and names, which require additional positional and ordering checks. In these cases, Dataguise uses a grammar-based approach to classify a group of candidate matches.

Finally, some elements are ambiguous. Although Dataguise produces no false negatives, it must remove all ambiguities to make an accurate determination of the sensitive data type and ensure zero false positives as well. Here, Dataguise performs contextual analysis, taking advantage of machine learning to get smarter. For example, a Visa credit card CVV (Card Verification Value) is a 3-digit number, and if all 3-digit numbers were classified as CVVs, certainly there would be many false positives. So Dataguise looks for credit card numbers, certain keywords, and other qualifying information in the vicinity of 3-digit numbers, before a CVV classification can be made.

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

3

Powerful, Proven Scalability to Grow with Your Business

The scale at which Dataguise Discovery can scan data repositories for sensitive data is unparalleled, whether scanning an entire repository or partial/sample data sets. It is uniquely optimized to leverage multi-threading and concurrent processing, whether deployed in a centralized or highly distributed infrastructure. Dataguise Discovery has been proven at some of the world's largest infrastructure deployments, including a 26-petabyte Teradata system at eBay. Even in the unstructured world of Hadoop, a typical Dataguise scan can be in the order of hundreds of terabytes, in which multiple scans are done, adding up to tens of petabytes—and this is done in a matter of minutes or hours, not months.

Real-World IoT Case Study:

EUROPEAN AUTOMOBILE MANUFACTURER

- New automobiles are fitted with over 40 microprocessors and dozens of sensors that monitor and collect telematics information (e.g., tire pressure, speeds, routes, wear and tear on components).
- An estimated 25 GB of telematics data are uploaded to the cloud per vehicle per hour to track the health of the vehicle and provide predictive alerts and maintenance.
- All of this data is tied to a Vehicle Identification Number (VIN), which is tied to the vehicle owner's identity and personal activity.
- Dataguise Discovery identifies the VINs in unstructured data streams and passes them to an integrated masking solution (also from Dataguise) to de-identify the owner.

4

Enterprise-Grade Quality, Flexibility, and Support for Added Confidence

Dataguise Discovery has been put through rigorous penetration testing and security certifications in large Fortune 500 companies around the globe. Its high-availability deployment architecture is designed for zero downtime, with 24/7 customer support and full disaster recovery. Licensing costs and models for Dataguise Discovery are flexible to meet the operational requirements of organizations of any size. Delivering an unlimited number of scans, it is extremely cost-effective and supports both manual and fully automated processes using command line or RESTful APIs.

DATAGUISE

5

Broad Platform Support for the Most Comprehensive Coverage

Dataguise Discovery can be deployed in siloed groups within an enterprise or at an enterprise-wide level, which is recommended and considered best practice for governance and compliance. It can find and report sensitive data in all flavors of Hadoop, relational databases, on-premises file shares, and object-based cloud stores. There is no other solution that can provide the breadth of platform-scanning capability at the element level that Dataguise delivers—and without the need to write a single line of code.

WHAT DATA REPOSITORIES CAN DATAGUISE SCAN?

(Representative list—contact us for more)

Relational databases (RDBMS) and structured data stores	Amazon Aurora Greenplum IBM Db2 IBM Db2 for z/OS IBM Power Systems (AS/400) MariaDB	Microsoft SQL Server MySQL Oracle PostgreSQL Sybase
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 HDInsight
NoSQL databases	Cassandra	
Cloud object stores	Amazon Simple Storage Service (S3) Google Cloud Storage Microsoft Azure Blob storage Microsoft Azure Data Lake Store	
Inflight data transfers	Apache Flume Apache Sqoop File Transfer Protocol (FTP)	
On-premises file servers	Linux Microsoft Windows	

DATAGUISE

WHAT'S DIFFERENT ABOUT DATAGUISE DISCOVERY?

The Dataguise software platform is the embodiment of more than a decade's worth of subject matter expertise in sensitive data discovery, protection, and monitoring—and in the most complex of environments that involve big data, IoT, and the cloud. Some enterprises have tried to use other security products or even built scripts in-house to find sensitive data accurately and thoroughly, but ultimately they chose Dataguise Discovery. There simply is no other solution that can do what Dataguise does. In the majority of situations, Dataguise will integrate with and augment any other solutions an enterprise already has in place for access management or perimeter-, network-, or data-level security.

HOW DOES DATAGUISE DISCOVERY COMPARE?

(Representative list—contact us for more)

Identity and access management solutions	These solutions are typically... • Not capable of finding personal data accurately, especially in unstructured data • Missing pseudonymization (masking) or encryption capabilities • Complementary to Dataguise
Encryption and tokenization solutions	These solutions are typically... • Limited to data protection, offering no discovery or monitoring capabilities • Using manual or assumptive methods to identify sensitive data, leading to false positives and false negatives, and a false sense of security and compliance • Not very useful for early breach detection and notification • Complementary to Dataguise
Data loss prevention (DLP) or network monitoring solutions	These solutions are typically... • Slow and difficult to scale • Unable to identify personal data in a table or file, especially in unstructured or free-form text • Able to offer some level of protection, but even pseudonymization is not static and can ultimately reveal original sensitive data • Complementary to Dataguise
Suite solutions claiming to do it all	These solutions are typically... • Immature offerings of classification, pseudonymization, and monitoring • Not proven in production at large-scale enterprises around the world • Limited in breadth of support for different data repositories
In-house, custom-built discovery solutions	These solutions are typically... • Very time-consuming and resource-intensive, especially without previous expertise • Usually built for a specific use case (data type or repository) only and require rework or result in limited scalability when additional data types or repositories are introduced

In one real-world case study at a major financial institution, a data loss prevention (DLP) solution took twice as long to find a fraction of the sensitive data that Dataguise discovered. Also, because DLP solutions usually perform at endpoints, there is a chance that sensitive data may have already been breached by the time it is discovered. Dataguise Discovery goes straight to the data source and performs at the data level—on demand or in near real time, as data is created or enters the enterprise—to discover, classify, and report the exact type and location of sensitive data. With Dataguise, enterprise stakeholders have complete and accurate visibility to manage risk and compliance with confidence.

> *To protect data, you must first know where users have stored it. Unfortunately, data has proliferated throughout the enterprise and can be difficult to discover. This is one of the significant struggles when security professionals attempt to deploy a DLP technology — if you can't locate where the enterprise stores its sensitive information, you don't know where to deploy controls.*
>
> — Forrester Research
>
> *The Future Of Data Security And Privacy: Growth And Competitive Differentiation*
>
> October 25, 2017

Part of the Dataguise Software Platform

How can you protect what you can't see? Sensitive data discovery is the first step in any data privacy, security or risk management strategy. That's why Dataguise Discovery is a foundational part of the Dataguise suite of software solutions, which also includes:

- Data Masking & Encryption: Safeguard data privacy and confidentiality in the event of an audit or breach with military-grade encryption and masking.
- Data Monitoring: Get real-time alerts of unauthorized or anomalous access behavior, and reduce the time it takes to identify a data breach from months to minutes.
- Data Auditing: Generate detailed compliance reports on what sensitive data is protected or exposed and who can access it.
- Executive Dashboard: Get a single, comprehensive view of sensitive data governance across the enterprise, including exposure, coverage, and trends.

All modules are seamlessly integrated to provide ultimate control and visibility of sensitive data. With Dataguise, executives get the answers they need to manage data breach and compliance risks. Innovators and decision makers get access to more and complete data sets for greater competitive advantage. Development and analytics teams get the freedom to leverage the power and efficiency of the cloud for accelerated ROI.

DATAGUISE

About Dataguise

Dataguise helps organizations deserve digital trust by protecting privacy as they use sensitive data to drive business value. The Dataguise software platform detects, protects, and monitors sensitive data as it is being collected, used, shared, and stored across the enterprise. With Dataguise, organizations can spend less resources on managing privacy, security, and risk and put more focus on using data to improve the products, services, and customer experiences that fuel business growth. To learn more, visit: http://www.dataguise.com.

SEE FOR YOURSELF–START A FREE TRIAL TODAY

Do your own proof of concept on premises or in the cloud. Dataguise is available now in the AWS Marketplace and in the Microsoft Azure Marketplace. For a limited time, qualified applicants can receive free credits to run Dataguise in the cloud at no cost. Just visit **http://www.dataguise.com/free-trial** to get started. Or call us any time at **877 632 0522 (U.S. toll-free)** or **+1 510 896 3770** to run an on-premises proof of concept at your company.