minereye
See Beyond Data

Fight Fire with Fire when Protecting Sensitive Data

White paper by Yaniv Avidan | January 2019

In an era where both routine and non-routine tasks are automated (diagnostic capsule detecting gastro-intestinal disorders, autonomous cars sensing their surroundings, home appliances controlled from your smart-phone to name just a few), Information Security still relies on the human factor, despite the inherent risk of errors or malicious actions.

It is time to put MinerEye's technology into use – the technology that automates all the tedious actions that require human intervention and governs your information assets.

Tomorrow's Battles

Modern warfare is shifting towards Cybersecurity. We are witnessing more and more sophisticated attacks from unexpected malware, or insider threats that force us to act fast and change our defense methodology accordingly. Individuals' and companies' sensitivity to data privacy is on the rise, while constantly trying to monetize on data, reduce risk and costs around unstructured data.

> "Unstructured data repositories have been chronically under-managed and overexposed within organizations, and the progressive adoption of cloud storage and collaboration platforms in recent years has made the situation even more complex to manage. A wave of new privacy regulations and continued threats such as crypto-ransomware are sending many organizations scrambling to find where sensitive data currently resides in their extended environment and who has access to it. This should not be approached as a onetime exercise, but rather as a call to institute sustainable and auditable processes to manage access to data."
>
> Gartner, Marc-Antoine Meunier, Published: 12 April 2017

PROPRIETARY INFORMATION

MinerEye's unique Interpretive AI™ is a proprietary Artificial Intelligence technology that addresses the challenges around undermanaged data repositories. It outputs actionable reports on sensitive documents and files on local or cloud repositories, helping companies to automate data protection policies and securely migrate data to the cloud. The combination of pattern recognition and machine learning automates the steps which were traditionally handled by sampled manual processes:

1) Continuously indexing and categorizing files into a searchable taxonomy;

2) Classifying data based on content and context;

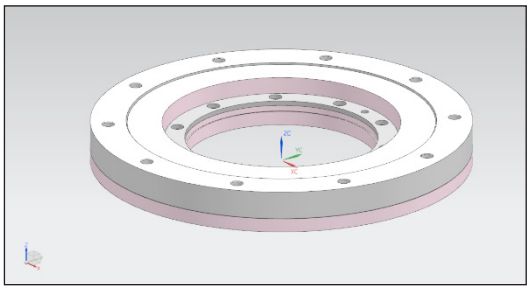
3) Selectively triggering security controls or workflows to minimize risk to your data.

These three vital steps will assist companies in getting ready for tomorrow's battles by buying time until a reaction and response to a breach will take place. As data becomes the new end point, the newest cyber threat will be data manipulation:

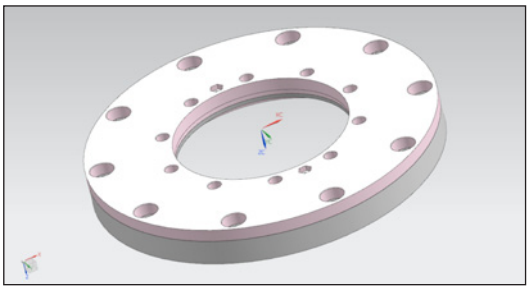
> "A 'Cyber Armageddon', long imagined in Washington as a catastrophic event of digitally triggered damage to physical infrastructure, is less likely than cyber operations that will change or manipulate data".
>
> US director of national intelligence, James Clapper at the House intelligence committee, Sep-10, 2015.

Learn and Scale

A fundamental component in automating data security is understanding your organization's full digital footprint without relying on static parameters. You can't secure what you can't see, and you can't control information assets that are in constant flux. Dynamic interpretive data discovery and categorization is the only solution for protecting the entire 'moving target' called the organization's digital footprint. In the vulnerable conditions we are all facing, scalability is the only answer. MinerEye's ability to scale in terms of data volumes, file formats, and constant changes both in content and in context is its distinct asset.



Learned



Matched

PROPRIETARY INFORMATION

Our scalability allows us the right flexibility to work with huge data volumes that are undergoing constant changes. By storing the data history and continuously looking for new and modified data incrementally, by working on the deltas (relative snapshot comparison mode), MinerEye's solution continuously learns, matches and updates the digital footprint of the company, and catches up on newly coming and changes around sensitive data. The solution dynamically scales across all data locations, and integrates with existing data security systems to ensure timely and accurate response of security controls.

Data profiling

Data profiling techniques are mostly utilized in the industry to make sure the data in structured databases is consistent, accurate, and reliable to support business activity. However, to make the profiling process agile and to cover all the unstructured and unknown data in all databases, is an even bigger challenge than the common profiling process. The data profiling process comprises structure discovery, data discovery and relationship discovery. Data profiling is performed using a tool that:

a) automates the discovery process; b) helps uncover the characteristics of the data; c) helps uncover the relationships between data elements in terms of format and type.

Traditional data profiling techniques face a substantial barrier when it comes to unstructured data, such as files, documents, images, computer aided designs (CAD), programs, etc. Structure discovery sometimes lacks reference: an employment agreement in one company has a different structure than another; insurance policy agreements vary between insurers; same for CAD methodologies of a next generation processing units, or any other binary-based intellectual property.

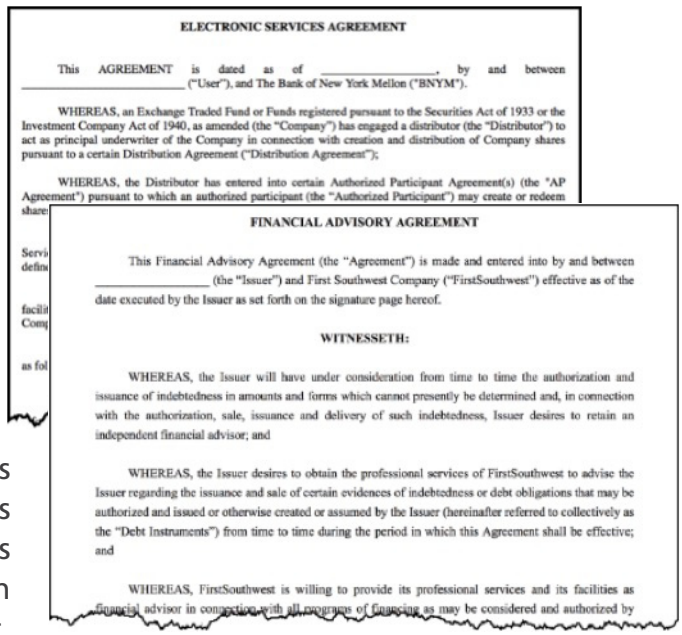
The solution lies in a complete conceptual change – to profile the data. MinerEye's DataTracker™ analyzes an exemplar file out of the organization's data set and applies pattern recognition and artificial intelligence techniques to perform structure discovery, data discovery and relationship discovery. Data Tracker™ contains algorithms that explore your data over multiple platforms and file types, retrieving the most accurate results, ordered by the level of similarity. For example, if you want to locate and track all the personal data that is stored on your file servers, you need to introduce a file containing such data for the system to train it to find all the similar files containing similar data elements.

After files with similar data are retrieved, you can choose to perform further actions, including:

- Automatically label and protect the data
- Segregate data according to location and restrict access through other locations

PROPRIETARY INFORMATION

- Minimize the data and monitor it, so that only relevant data is stored on the right locations
- Automate data categorization
- Perform focused forensic analysis



ELECTRONIC SERVICES AGREEMENT

This AGREEMENT is dated as of _____, by and between _____ ("User"), and The Bank of New York Mellon ("BNYM").

WHEREAS, an Exchange Traded Fund or Funds registered pursuant to the Securities Act of 1933 or the Investment Company Act of 1940, as amended (the "Company") has engaged a distributor (the "Distributor") to act as principal underwriter of the Company in connection with creation and distribution of Company shares pursuant to a certain Distribution Agreement ("Distribution Agreement");

WHEREAS, the Distributor has entered into certain Authorized Participant Agreement(s) (the "AP Agreement") pursuant to which an authorized participant (the "Authorized Participant") may create or redeem share

FINANCIAL ADVISORY AGREEMENT

This Financial Advisory Agreement (the "Agreement") is made and entered into by and between _____ (the "Issuer") and First Southwest Company ("FirstSouthwest") effective as of the date executed by the Issuer as set forth on the signature page hereof.

WITNESSETH:

WHEREAS, the Issuer will have under consideration from time to time the authorization and issuance of indebtedness in amounts and forms which cannot presently be determined and, in connection with the authorization, sale, issuance and delivery of such indebtedness, Issuer desires to retain an independent financial advisor; and

WHEREAS, the Issuer desires to obtain the professional services of FirstSouthwest to advise the Issuer regarding the issuance and sale of certain evidences of indebtedness or debt obligations that may be authorized and issued or otherwise created or assumed by the Issuer (hereinafter referred to collectively as the "Debt Instruments") from time to time during the period in which this Agreement shall be effective; and

WHEREAS, FirstSouthwest is willing to provide its professional services and its facilities as financial advisor in connection with all programs of financing as may be considered and authorized by

MinerEye has developed a proprietary technology to match data to a reference set using advanced techniques applied originally in computer vision tasks. These algorithms read the bytes of a given exemplar file, and represent its content by creating a *signal* – a single mathematical vector on the fly, weighing only a few K's per file. From this point onward, Data Tracker™ uses all the signals to create signal clusters and performs other analysis tasks. This method allows the system to process unprecedented amounts of information about the originally scanned data set. The system does not move files from their original storage location, neither does it open or change them, but rather reads their byte stream, creates the signals and sends them out to the server. This saves enormous network load while learning the sensitive data patterns and attributes. The signal can be refreshed on a scheduled basis since the data doesn't change at a pace that affects the accuracy of the clustering process. The scheduled scans are obviously incremental.

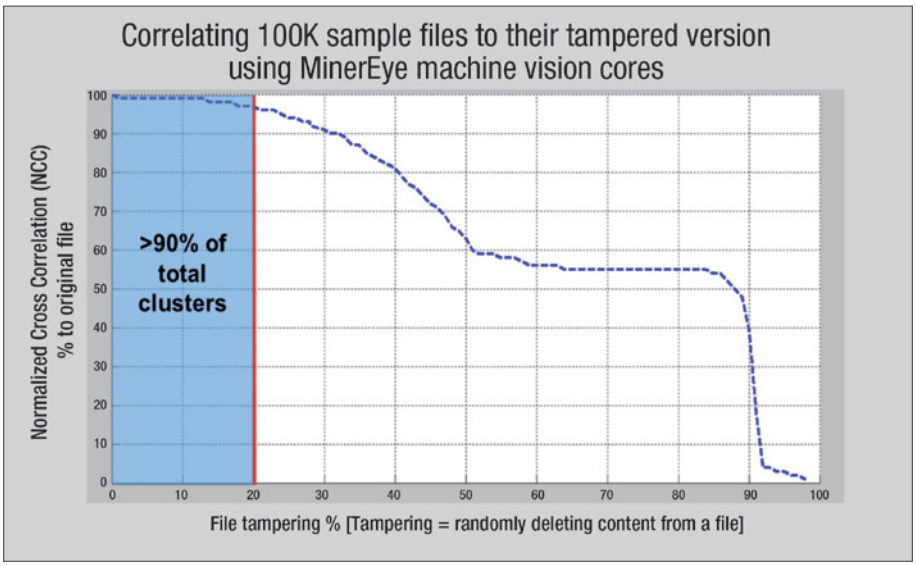
Data protection is based upon learning to dynamically identify it and understand its behavior. When implementing a data protection program, CISOs face the challenge of defining the start and end points of such program. Current solutions and technology constraints compel the organization to spend huge amounts of money and resources to define and identify sensitive data.

Another challenge CISOs are facing is the volumes of data and the rate it is being created, changed and multiplied. When trying to overcome this issue, one of the assumptions that can be made is that most of the data within the organization is duplicated with slight changes to content, throughout all business units and business processes. There is a lot of reuse taking place when generating reports, presentations, code files, computer aided designs, and so on; just think how many times the *Save As* operation is performed in a company each working day. This means that most of the data can be grouped into clusters of similar versions.

PROPRIETARY INFORMATION

MinerEye's Data Science team proved this assumption while running analysis algorithms over large and heterogeneous data sets taken from several types of organizations. The results showed that even files whose content has been physically changed by 20% were matched by the system to their original version by a very high confidence using the Normalized Cross Corrlation metric (NCC >90%). The results also showed that over 90% of total clusters generated by the system are located in the same area of >95% correlation. This data characteristic is the key to turning the data categorization process into a manageable one compared to any other method.

Transforming the flat structure of data as it is distributed across storage locations, shares and desktops into a hierarchy of similarity clusters, provides a searchable taxonomy that can be easily visualized and enables the analysis of huge amounts of data. Additionally, such scalability is required when dealing with large amounts of data streaming into the scanned and clustered repositories.



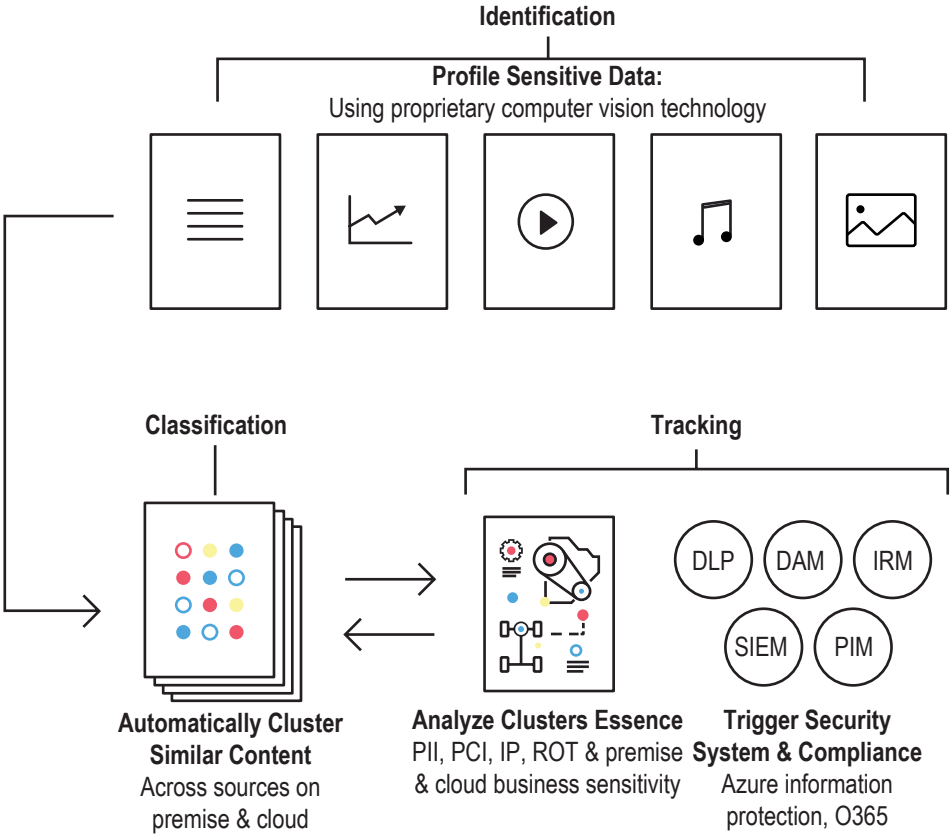
Fight Fire with Fire

MinerEye's Data Tracker™ is a virtual machine that encapsulates all the technology that we have developed and mastered – AI-driven automated data protection and migration, AI-driven classification and categorization, and continuous data profiling techniques, followed by automatic triggering of security instruments.

Data Tracker™ offers a three-step funneled process to establish the foundation for optimized data protection and for profiling sensitive data and triggering the relevant action.

Profiling the data is the basic step in the deployment of the automated classification process. During the initial scan, the system crawls the entire database in all repositories and creates signals for every file, regardless of its type, and sends these signals to MinerEyes' server.

PROPRIETARY INFORMATION

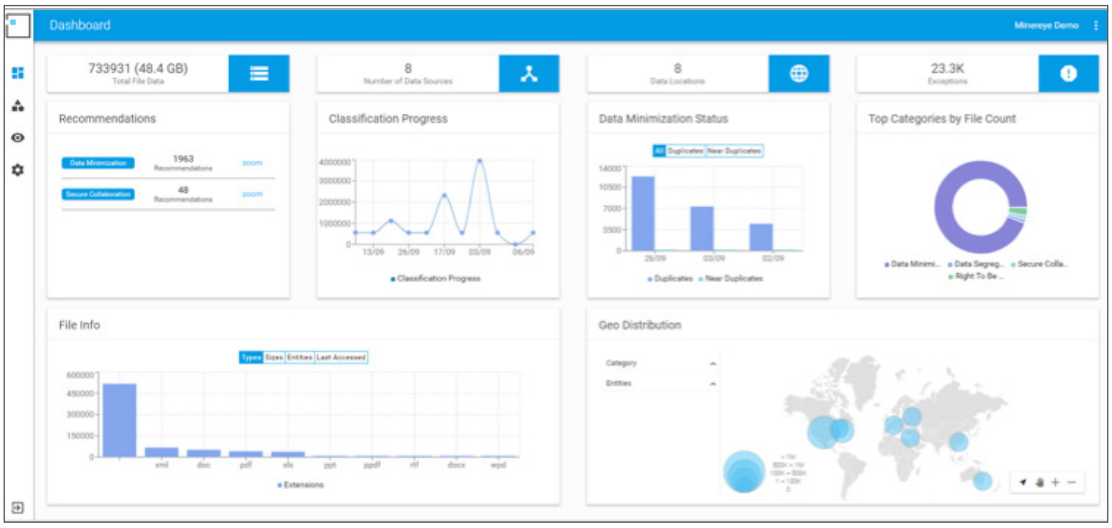


Classifying the data is the second stage of the process that enables the tracking of data in a specific context according to the organization's data classification conventions. This step includes clustering the data and matching examples of sensitive data elements to existing clusters, in preparation for the tracking process that will follow. These steps play a vital role in indentifying exisiting and new incoming data.

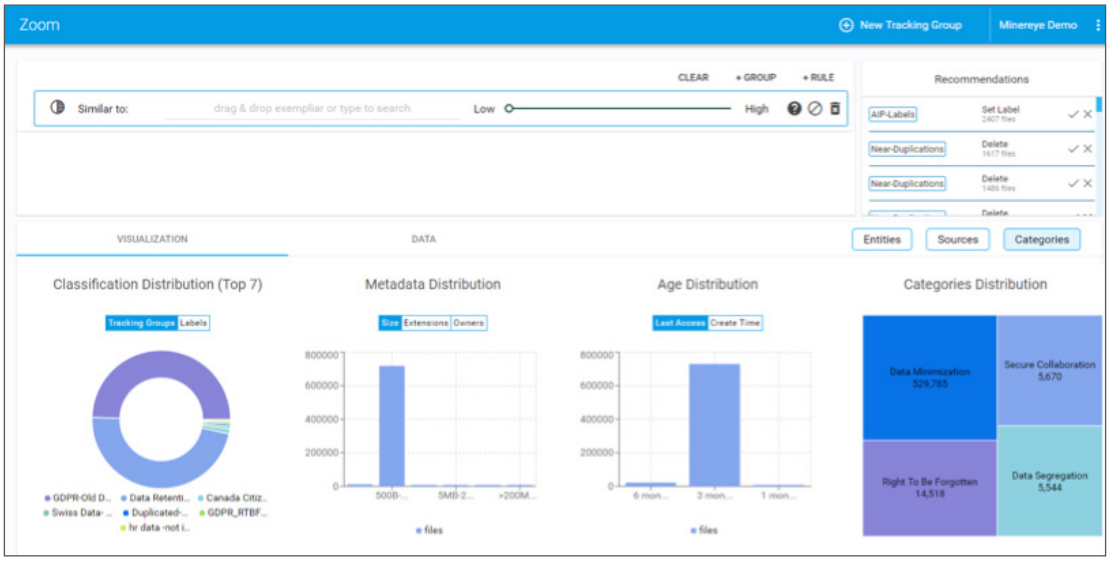
Tracking the data When a new file is identified during the system's incremental scan (every scan that is performed after the initial scan), it is automatically matched to a cluster. The system can apply a contextual rule to each cluster to enable tracking files in a specific context, e.g. files in the cluster that have been identified out of a specific folder or geography, or files that have been last modified before a specific date and are aging. The user can then tag the cluster with a meaningful name or by applying the organization's classification convention.

PROPRIETARY INFORMATION

DataTracker's Dashboard provides the user with an instant view of the organizational data. It provides a summary of the organizational data according to various statistics along with the Handling Recommendations panel, which provides actionable recommendations by categories.



Data Tracker's GUI enables the user to build rules that allow him to perform a complex search for a file, find all its similars, and filter the results according to desired entities, file types, metadata, API labels, and more.



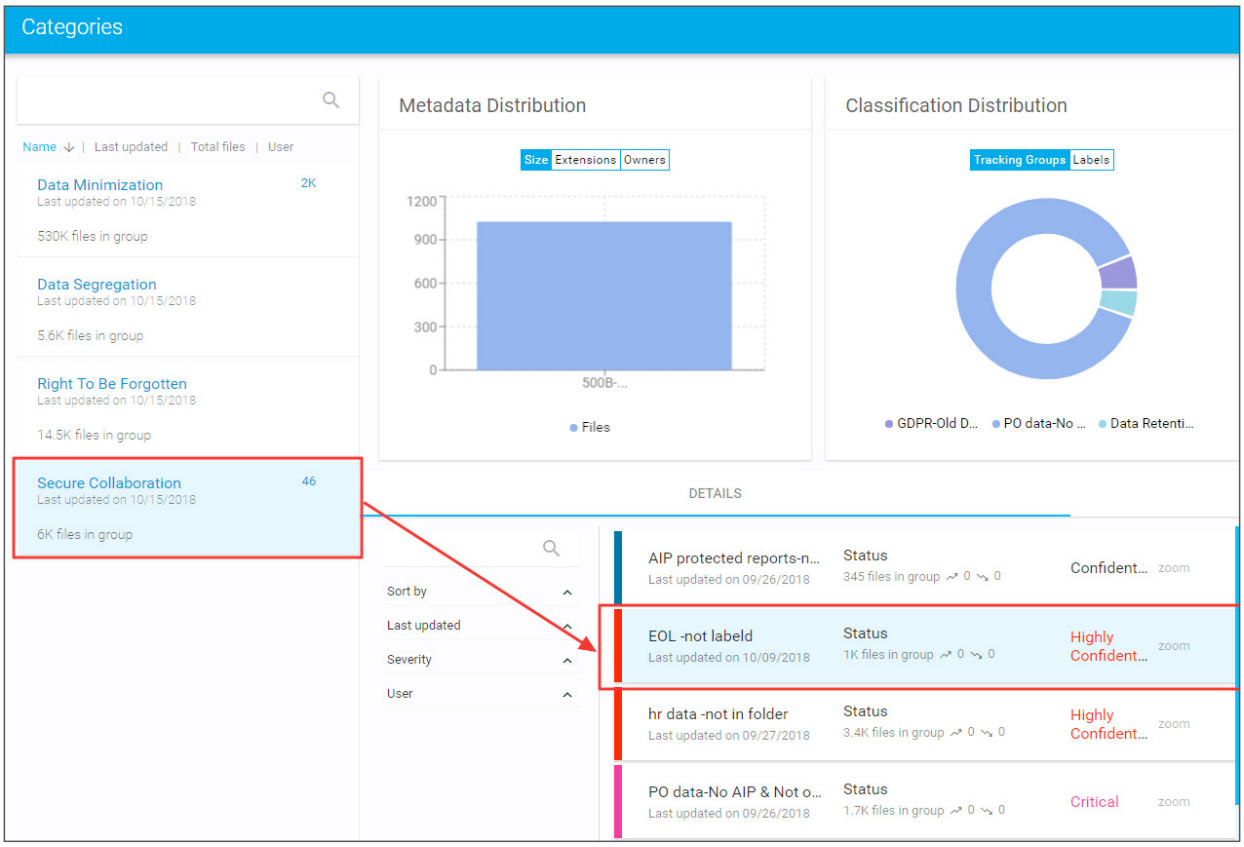
Triggering systems to action is the output of the continuous profiling process that converges two abilities of the Data Tracker™ platform:

a) A set of API's interacting with external security controls;

b) An Artificial Intelligence module that profiles the behavior of sensitive data clusters and reports on their locations.

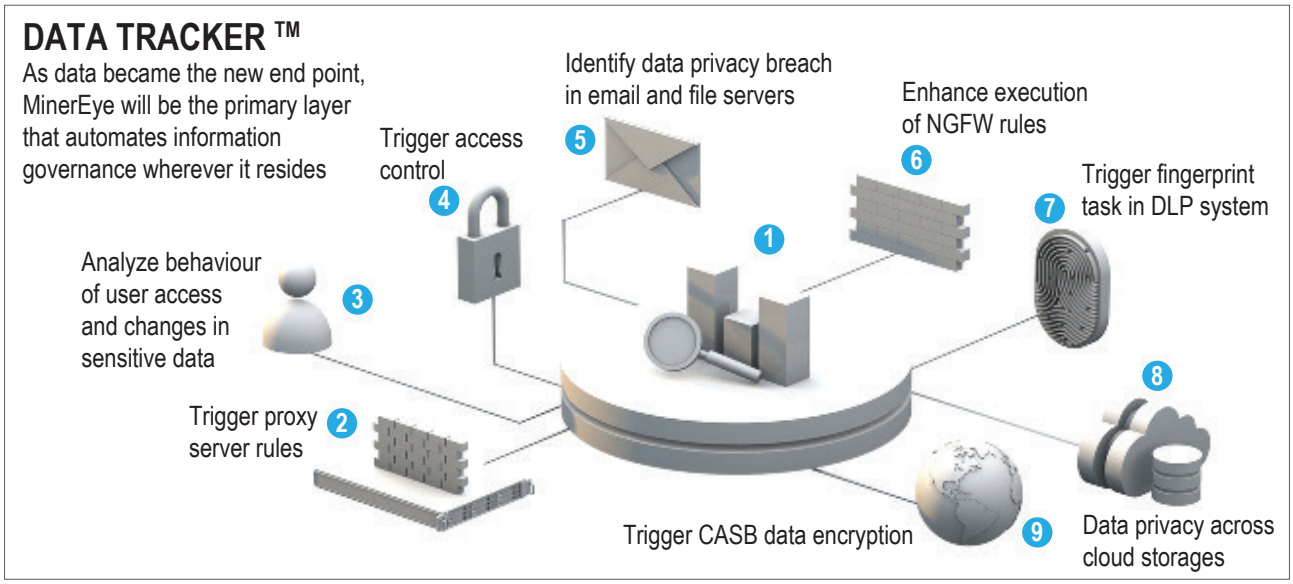
PROPRIETARY INFORMATION

DataTracker's Categories module allows the user to maintain Tracking Groups that alert the user on security breaches and can trigger Data protection or management systems to act accordingly.



Most DLP and IAM systems are configured to protect several folders that contain sensitive data. These systems apply a pre-configured policy over identical files found by their discovery process in external systems such as endpoints, shares, email and other proxy servers. Some even enhance the policy enforcement task by searching on regular expressions defined in their internal dictionary. Cloud access brokers apply access permissions or application lock based on predefined parameters that would imply on the nature of the data that is being shared or dispositioned. Naturally, in a hyper-dynamic data environment, this would not be sufficient to cover all data from volume and type perspectives. Data Tracker™ utilizes its APIs to improve the security controls' throughput. The system uses the protected folders of DLP and IAM systems as learning sets to discover new locations of similar data that couldn't be discovered by any other system, and automatically triggers a fingerprinting task or an access control policy within these systems over those new locations. Data Tracker™ can trigger an encryption or application lock command within a Cloud access security broker (CASB) upon identification

PROPRIETARY INFORMATION

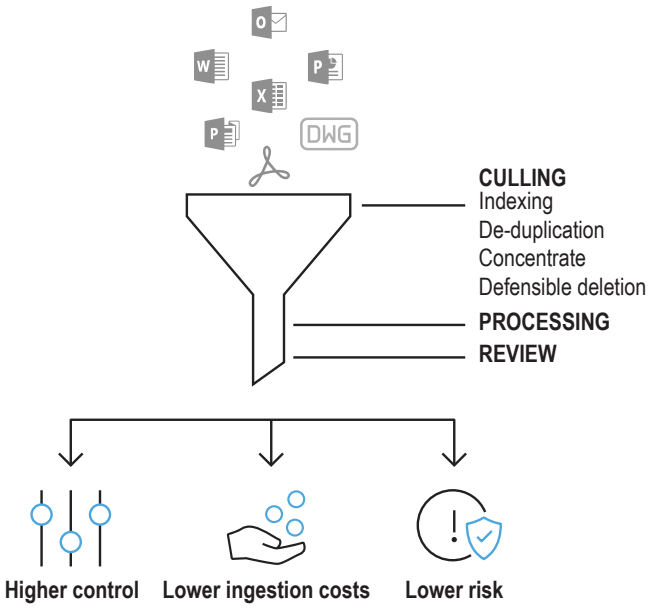


of data that couldn't be identified by the CASB. The same goes for firewall rule set and proxy policy. Data Tracker™ enriches SIEM and SOC platforms with outlier incidents that are sensitive data-focused, and provides a holistic view of risk to the organizational data. Our APIs enable the querying of the database for forensic purposes, by analyzing the history of confidential files, and supporting data breach investigations.

Use Cases

Data minimization

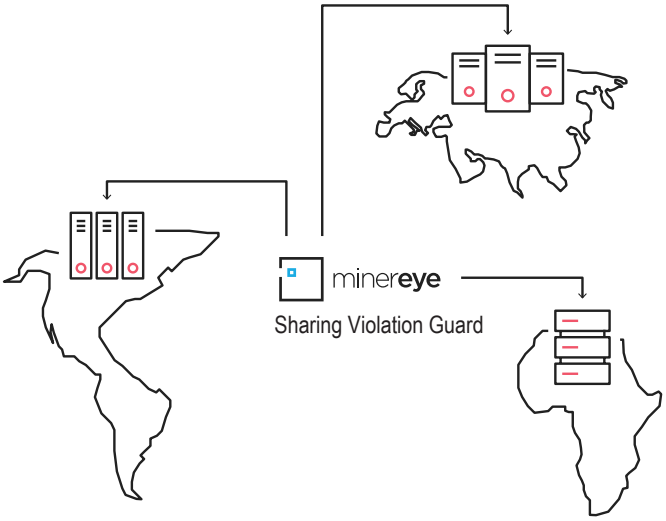
MinerEye helps companies get rid of obsolete and redundant data that has accumulated over time and takes up precious data storage. MinerEye technology couples with methodologies and experts, allowing you to overcome common data minimization challenges, such as reconciling conflicting regulatory requirements, dealing with data packrats, battling zombie data and getting the "SharePoint menace" under control.



With MinerEye's efficient data minimization capabilities, companies can save up to 30% of their storage space, while ensuring compliance with regulatory requirements.

PROPRIETARY INFORMATION

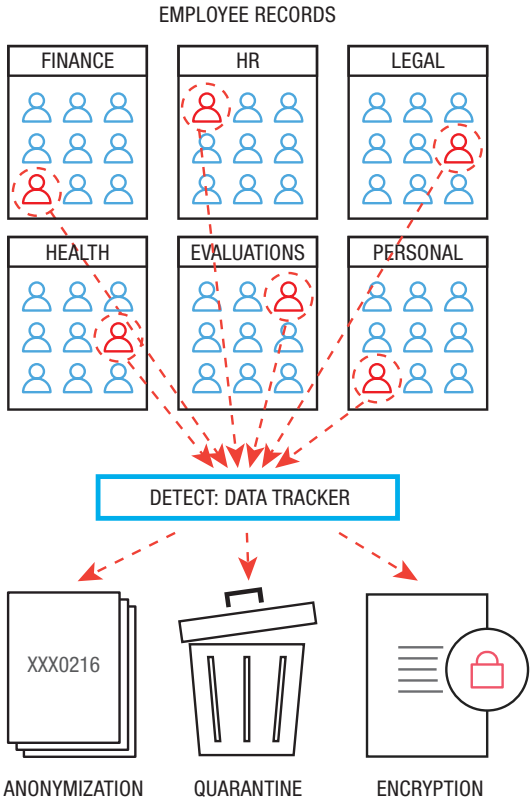
Data segregation



MinerEye's AI continuously maps and alerts on data segregation issues and data sharing violations that are prevalent among large dataset carriers. Companies subject to regulations are required to classify documents within their endless sets of data and segregate them into restricted locations, while assuring the correct storage location of all files. Data is not static – it is constantly created, modified, duplicated, and incorporated into other document forms, causing endless complexities, especially for global organizations. MinerEye's AI speed outpaces the overwhelming rate of data creation, modification and movement, identifying unregulated transferal of data across geographic boundaries, guaranteeing that all data is in complete order and control at all times. MinerEye's solution architects help companies deploy a net of virtual machines (VM) across company sites, to ensure compliance with data residency and sovereignty regulation.

The right to be forgotten

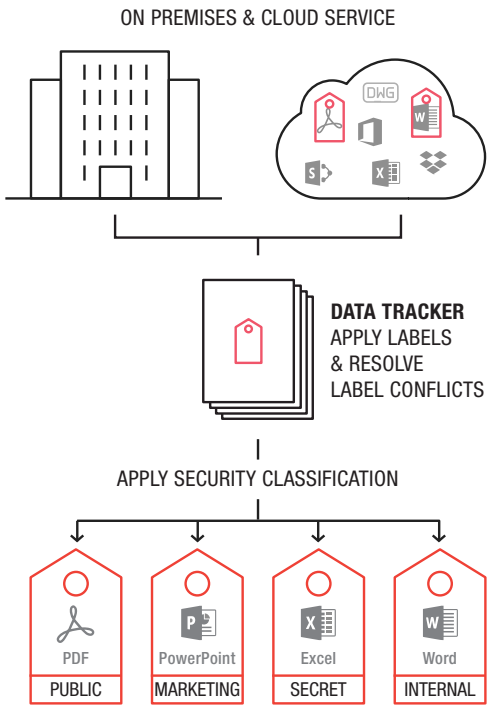
The right to be forgotten category allows the user to handle situations when customers, vendors, partners, or employees request to find and eliminate any data connected with them. DataTracker enables the tracking of unstructured data that is related to personal data elements. Upon discovery, the system triggers an external process to delete the required data.



PROPRIETARY INFORMATION

Secure collaboration

Data collaboration between users in the organizational cloud service or platform is often vital, but business processes often move the data across platforms, such as email, collaboration applications, and so on; thus, creating an immediate risk of exposure, leakage and lack of control over the data. To prevent such exposure, DataTracker offers automated and contextual classification of files. The integration of DataTracker with Azure Information Protection (O365) labeling mechanism can enforce file labeling and automate the propagation of that label or policy onto similar content across platforms, using DataTracker AI capabilities.



DataTracker detects unlabeled files or labeling conflicts and recommends the correct labeling regardless of the platform they're stored on. The system can detect discrepancies between labels, namely, files that are similar in content, but are labeled differently or not labeled at all and recommend to you what labels should be applied.

The Secure Collaboration use case ensures that the organizational security policy is properly implemented, and that data can be safely collaborated over the cloud infrastructure.

PROPRIETARY INFORMATION

minereye
See Beyond Data

MinerEye is the pioneer in Interpretive-AI for Secured Cloud Migration. Employing advanced computer vision and machine learning technologies, MinerEye offers the most comprehensive vehicle for identifying and tracking organizations' sensitive data, anywhere across the globe and throughout the unstructured scales of "dark data."

MinerEye AI technology enables secure cloud adoption for the world's leading enterprises, including leading banks and technology vendors.

Contact Us

telephone: +972-9-7653712

email: info@minereye.com

Visit our website: MinerEye.com