

## A TECH GUIDE

---

# Data Discovery



# CONTENTS

<b>Data Discovery is Business Critical</b>	<b>3</b>
<b>Planning for Effective Data Discovery</b>	<b>5</b>
<b>GLASS Technology™ for Dependable Discovery</b>	<b>7</b>
<b>Data Discovery is an Opportunity</b>	<b>9</b>
<b>Contact</b>	<b>10</b>

# Data Discovery is Business Critical

Organizations depend on the information they gather for market intelligence and as a profit-generating commodity. Across the world, the information economy is booming. The IDC predicts that stored data worldwide will reach 175 zettabytes — 175 trillion gigabytes — by 2025. If stored on Blu-ray, that's enough data to reach the moon 23 times over.<sup>1</sup>

As data continues to proliferate, it can be difficult to keep track of everywhere it resides. As the availability of computing resources increases, so too do the potential locations for data to be stored. Migration to cloud and virtualized infrastructure only makes this more challenging.<sup>2</sup>

Organizations need to make data awareness a fundamental part of any data management strategy. Data awareness starts with data discovery: understanding the data, where it is, how it is used and what measures are in place to protect it.

## Data Discovery and Compliance

Much of this deluge of data is Personally Identifiable Information (PII) or sensitive information. With over 130 countries now enforcing data protection legislation to protect PII and consumer privacy,<sup>3</sup> organizations are increasingly struggling to meet their obligations.

Gartner predicts that by 2023, 65% of the global population will have their PII covered by privacy regulations,<sup>4</sup> all of which enforce penalties for poor data management that results in data breach or misuse of data. According to the site Privacy Affairs, GDPR breaches have resulted in fines exceeding €1.6 billion<sup>5</sup> since its introduction in May 2018.

In addition to the regulatory burden, organizations are also required to satisfy other standards such as PCI DSS, enforced through contracts and subject to similar financial penalties for non-compliance and data breach.<sup>6</sup>

By 2025 stored data worldwide will reach

**127**  
trillion  
gigabytes



**Over 130**  
countries now  
enforce data  
protection  
legislation

## Built on Trust

Data on its own isn't the profit driver for most organizations, but it enables organizations to nurture their most valuable assets — their customers. The trust and loyalty of customers is crucial to success for most businesses. While building and maintaining customer trust can be a long and time-consuming process, losing it is instant. When Facebook was embroiled in the 2016 Cambridge Analytica scandal their share price fell by over \$130 billion (since recovered).<sup>7</sup>

One 2022 study from The Harris Poll,<sup>8</sup> found that 75% of consumers in the US and UK are uncomfortable sharing their data with companies that are known to have “poor data ethics”. Nearly all respondents (89%) noted that they would be more likely to spend their money with a brand that commits itself to protecting their personal data online.

Data breaches can also come with immediate market devaluations. When it was revealed in 2016 that Yahoo had suffered two of the largest breaches in history, amid an acquisition deal with Verizon, Yahoo was forced to close at a \$350 million discount.<sup>9</sup>

## Driving Value from Data

Since 2019, the number of organizations recruiting chief data officers (CDOs) has doubled. Over 20% of the top 2,500 publicly traded companies have a CDO in place,<sup>10</sup> establishing data management and security as a critical C-Suite function. Alongside responsibility for data security and data management, the CDO and equivalent roles aim to maximize value from the data held by their organization. With data awareness so fundamental to an effective data strategy, CDOs are dependent on understanding the nature and whereabouts of their data to deliver their objectives.



CDOs are dependent on understanding the nature and whereabouts of their data to deliver their objectives.



**Facebook share price fell**

**\$130 billion**

**after Cambridge Analytica scandal**

**89%**

**of people more likely to spend money with a brand that is protecting their personal data**

# Planning for Effective Data Discovery



There are three main components to planning an effective data discovery exercise:

- 1 > Defining its purpose and scope
- 2 > Selecting the discovery tools or solutions to be used
- 3 > Analyzing the results and remediating any findings

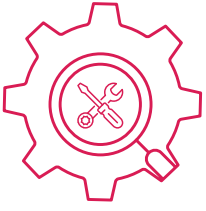


## 1. Understanding Why and What for Efficient Scanning

The extent of a data discovery exercise depends on the goals for the scan. For effective scanning, tools need to be configured for the type of data to be identified in all the possible forms and formats it may be recorded. Depending on the purpose of the scan, the exercise may comprise network discovery to identify all possible scan targets, or concentrate on specific network segments or components.

For example, a data discovery scan to establish scope for a PCI DSS compliance program will need to cover the entire network to ensure that all possible stores of account data are identified.<sup>11</sup> Whereas a discovery scan to confirm that customer records have been purged from a database can target that system in isolation.

Advanced discovery solutions will support network discovery, which identifies targets located within a defined IP address range for scanning during data discovery.



## 2. Choosing the Right Tools for the Job

Many organizations run scripts or RegExes to sample their environments for PII and sensitive data. The problem with these methods is that the results are dependent on the quality of the script or query. Too often, these don't account for the various forms in which a data type may occur (for example, whether a card number occurs as an unbroken string of 16 digits, or as four blocks of four digits separated with a space or dash), exclude parts of the network or are incompatible with business systems. Prone to false positives and — even worse in PII terms — false negatives, data discovery using scripts and RegExes is not robust enough to provide reliable results for effective data management.

More advanced data discovery solutions, such as Ground Labs Enterprise Recon, provide out-of-the-box discovery capabilities for a far wider range of data types as well as the various forms in which they may be found. Enterprise Recon can identify more than 300 types of personal, sensitive and confidential data from over 50 countries, in predefined and variant forms.<sup>12</sup>

## 3. Interpreting Results and Planning Remediation



Once the discovery exercise is complete, the results need to be compiled for analysis. With script-based outputs, this can be an onerous manual process requiring the scrutiny of personnel with appropriate knowledge and experience of the business, its network, processes and systems to be able to highlight potential false or anomalous findings.

Purpose-built data discovery solutions offer a more comprehensive suite of options for reporting and deliver greater accuracy than script or search-based methods. Ground Labs products export results in a range of standard formats and can be integrated with several reporting suites including Tableau, Power Bi, Crystal Reports and MS Information Protection.<sup>13</sup>

Following analysis, any confirmed findings need to be addressed in line with the original purpose of the discovery exercise. This is typically a manual change process, but it can be simplified using solutions specifically designed for data discovery that offer remediation-in-place capabilities. These solutions facilitate effective data management by enabling organizations to address any rogue data quickly and efficiently, mitigating the associated risk.

# GLASS Technology™ for Dependable Discovery

Recognizing the limitations of many discovery methods, Ground Labs developed its proprietary GLASS Technology™ that is at the core of its Enterprise Recon data discovery solutions. Unlike scripted and RegEx-based discovery tools, advanced solutions such as Enterprise Recon provide data discovery out-of-the-box, preconfigured for a wide range of data types, varieties and forms in which data may be presented. Agnostic to the multitude of file types, platforms and network environments within an organization, Ground Labs data discovery products provide comprehensive, accurate and fast results.

## ENTERPRISE RECON



Enterprise Recon is powered by GLASS Technology, a pattern-matching engine that enables Ground Labs discovery solutions to **operate across multiple platforms and match against multiple patterns simultaneously**. Its custom pattern-matching algorithm is built on a succession of filters optimized for each CPU cache stage. This allows it to disregard data that doesn't meet its matching criteria early, ensuring time- and resource-efficient scanning.



**Accented characters are handled with automatic character normalization**, ensuring that all data can be incorporated in discovery scans. These may be excluded, ignored or omitted in other data discovery methodologies. Pattern-matching is further enhanced by the ability to customize matching of metadata providing even more granular control over the discovery process.



Many data discovery tools overlook the importance of context in establishing the sensitivity of data. **Ground Labs uses contextual data analysis** that further filters data based on other information stored alongside it, scored against a predetermined threshold. This capability means that data stored in association with other PII data can be flagged even if the specific data type may not otherwise be considered sensitive.

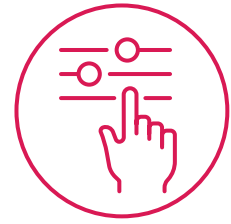


Beyond its powerful pattern-matching engine, **Ground Labs solutions use checksums, function calls and other methods for data validation**. For payment card details, for example, any details matching the criteria for a payment card number will be validated using the Luhn check.<sup>14</sup> These various validation methods help to prevent false positives from being identified and reported.

## Customization for Targeted Discovery

The pattern-matching technology underpinning Ground Labs Enterprise Recon solutions is flexible enough to support customization beyond the 300 data types configured in the tool.

Every business is different and the data that is most sensitive is equally varied. While Ground Labs focuses on PII and regulated data, Enterprise Recon supports data discovery for any other data through customization. The customization process, programmed via its easy-to-use interface, configures the GLASS Technology engine to search for any data of value to an organization.



## Remediation-in-Place and Post-Discovery Data Management

**The objective of most data discovery exercises is to support compliance with security standards and frameworks, to meet contractual, regulatory or legislative obligations. Audit-ready reporting is a feature that many organizations should consider when selecting a tool for data discovery.**

Enterprise Recon provides highly accurate reports in several formats and, for its PRO edition, offers integration with ODBC-compliant business intelligence tools for data analytics and advanced reporting.

Data discovery is only one part of the data management puzzle. Analyzing findings to determine an appropriate response is another key component in an effective data strategy. Dashboards that provide risk scoring against discovery scan findings means that remediation tasks can be prioritized effectively.

Implementing automated data discovery scanning — in a solution with remediation-in-place, data classification and data access management capabilities — as part of a wider data management strategy, ensures that unauthorized data stores or unexpected pockets of data are identified and addressed quickly.

With a clear dashboard offering a consolidated view of risk exposure, and with remediation options including masking, quarantine, encryption and deletion of data,<sup>15</sup> Ground Labs data discovery solutions facilitate efficient data management for organizations of every size.





# Data Discovery is an Opportunity

Data is too valuable and too sensitive a subject to be ignored. It's not just that poor data handling will lead to regulatory blowback, cyberattacks or reputational damage — it's a lost opportunity. Data is more sensitive than ever, but it's also more valuable than ever. By neglecting such a fundamental asset, opportunities for innovation, insight and profit are lost.

Using data discovery as the foundation of an effective data management strategy, in an automated business-as-usual process, provides organizations and CDOs greater visibility of their data, enabling them to protect it while maximizing its value.

Discovery solutions that provide direct remediation, classification and access management of PII and other sensitive data:

- reduce the risk exposure of the data
- protect consumer trust and
- support regulatory, legal and contractual compliance

## ENTERPRISE RECON

Ground Labs' data discovery solutions take the guesswork out of compliance. With scalable solutions to meet your needs, Ground Labs is the discovery partner you need.

Ground Labs Enterprise Recon is trusted by organizations for its comprehensive card data discovery and remediation-in-place capabilities.

Enterprise Recon provides a centralized view of all data stored across cloud and on-premises systems. It can identify over 300 types of data out of the box. Powered by Ground Labs proprietary and award-winning GLASS Technology™, Enterprise Recon is the quickest and most accurate discovery software on the market, enabling rapid data discovery, risk scoring and data classification and remediation-in-place across the broadest set of platforms.

[groundlabs.com/enterprise-recon](https://groundlabs.com/enterprise-recon) ►



Established in 2007 and trusted by more than 4,500 companies in 85 countries, Ground Labs offers award-winning data discovery and management solutions for all industry sectors.

[www.groundlabs.com](http://www.groundlabs.com)

**CONTACT:**

US **+1 737 212 8111**  
UK **+44 203 137 9898**  
Ireland **+353 1 903 9162**  
Australia **+612 8459 7092**  
Asia **+65 3133 3133**

---

Email **[info@groundlabs.com](mailto:info@groundlabs.com)**

# Endnotes

1 <https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html>

2 <https://www2.deloitte.com/xe/en/insights/industry/technology/technology-media-and-telecom-predictions/2021/cloud-migration-trends-and-forecast.html>

3 <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>

4 <https://www.gartner.com/en/newsroom/press-releases/2020-09-14-gartner-says-by-2023--65--of-the-world-s-population-w>

5 <https://www.privacyaffairs.com/gdpr-fines/>

6 <https://www.pcidssguide.com/what-are-the-pci-compliance-fines-and-penalties/>

7 <https://www.cbsnews.com/news/facebook-stock-price-recovers-all-134-billion-lost-in-after-cambridge-analytica-datascandal/>

8 <https://www.businesswire.com/news/home/20220720005515/en/New-Research-Finds-That-75-of-US-and-UK-Consumers-Are-Not-Comfortable-Purchasing-From-Brands-With-Poor-Data-Ethics>

9 <https://www.cbsnews.com/news/verizon-yahoo-merger-price-data-breaches/>

10 <https://www.strategyand.pwc.com/de/en/digital/cdo-2022.html>

11 [https://docs-prv.pcisecuritystandards.org/Guidance%20Document/PCI%20DSS%20General/Guidance-PCI-DSS-Scoping-and-Segmentation\\_v1\\_1.pdf](https://docs-prv.pcisecuritystandards.org/Guidance%20Document/PCI%20DSS%20General/Guidance-PCI-DSS-Scoping-and-Segmentation_v1_1.pdf)

12 <https://www.groundlabs.com/enterprise-recon/>

13 <https://www.groundlabs.com/documentation/er/Content/Analysis/Reports.html>

14 <https://www.groundlabs.com/blog/anatomy-of-a-credit-card-the-luhn-algorithm-explained/>

15 Remediation-in-place options vary depending on the data source, location and file type.

## COPYRIGHT NOTICE

© 2022 Ground Labs. All Rights Reserved. The Ground Labs name and logo and all other names, logos and slogans identifying Ground Labs products and services are trademarks and service marks or registered trademarks and service marks of Ground Labs Pte Ltd and its affiliates in Singapore and/or other countries. All other trademarks and service marks are the property of their respective owners.

DOCUMENT LAST UPDATED:  
**DECEMBER 2022**