A Practical Guide: The Data Mapping & System Inventory Framework

Michael Schneider

DIRECTOR AND PRINCIPAL SPECIALIST SOLUTION ARCHITECT, ZSCALER

WHITEPAPER

Table of Contents

Introduction: The Architecture of Visibility

In a modern zero-trust environment, “data is the new perimeter.” However, you cannot protect what you cannot see and data is no longer a static resource housed within the safety of a single data center. Instead, it has become a fluid asset that flows across multi-cloud environments, thousands of SaaS applications, and a global, mobile workforce. As organizations race to embrace Generative AI and broader digital transformation, the traditional perimeter-based security model has been rendered obsolete, leaving security leaders with a fragmented and often inaccurate view of their most critical assets. This document introduces a strategic, three-phase methodology designed to move organizations from a state of chaotic data sprawl to a disciplined posture of Content-Aware Zero Trust. This document outlines the rigorous process of building a Living Data Inventory—a dynamic record that connects business processes to technical assets and regulatory requirements.

The framework begins with the Visibility Pillar, which is focused on identifying every system that contains sensitive data. This initial phase is the foundation of the entire security strategy, as it is impossible to protect an asset that remains invisible to IT. By mapping the complete data estate, organizations can illuminate hidden cloud storage, unsanctioned “Shadow IT,” and forgotten legacy systems. This process moves the security team away from manual, static inventories and toward an automated, real-time discovery process that ensures every repository—whether managed or rogue—is accounted for and brought under the organization’s security umbrella.

Building upon that visibility, the second phase establishes the Governance Pillar by defining and verifying system ownership. Visibility alone is insufficient if there is no clear line of accountability for the data being discovered. This phase bridges the gap between the technical custodians who manage the infrastructure and the business stakeholders who understand the data’s true value and risk. By creating a rigorous lifecycle of ownership attestation, organizations can eliminate “orphaned” data and ensure that security policies are grounded in business reality rather than technical guesswork. This structural alignment ensures that risk is explicitly owned and that access is continuously justified by those most qualified to make that determination.

The methodology culminates in the Intelligence Pillar, which involves the automated discovery and classification of the data content itself. While the previous phases address the “where” and the “who,” this final step provides the “what.” By leveraging advanced AI and machine learning, organizations can look inside their systems to distinguish between public information and highly sensitive intellectual property or regulated personal data. This level of granular intelligence is the prerequisite for modern security initiatives, particularly the safe deployment of Generative AI, as it allows for surgical enforcement of Data Security policies that prevent leakage without obstructing legitimate business workflows.

The Zscaler Data Security Platform serves as the central engine for this entire methodology, providing a unified fabric that connects these three disparate phases into a single, cohesive workflow. Unlike legacy approaches that rely on a patchwork of disconnected point products, Zscaler offers a single pane of glass to inspect data at rest, in motion, and in use. By integrating these capabilities into the Zero Trust Exchange, Zscaler enables security leaders to automate the journey from data blindness to total intelligence. This document serves as a comprehensive guide for leveraging that platform to reduce the attack surface, achieve continuous compliance, and secure the organization’s most valuable assets in an increasingly complex digital world.

Part I: Identifying Systems Containing Sensitive Data

System identification is a multi-modal process. Relying on a single source (like a CMDB) often results in a 30–40% “blind spot” due to Shadow IT and rapid cloud adoption.

Why is this Needed?

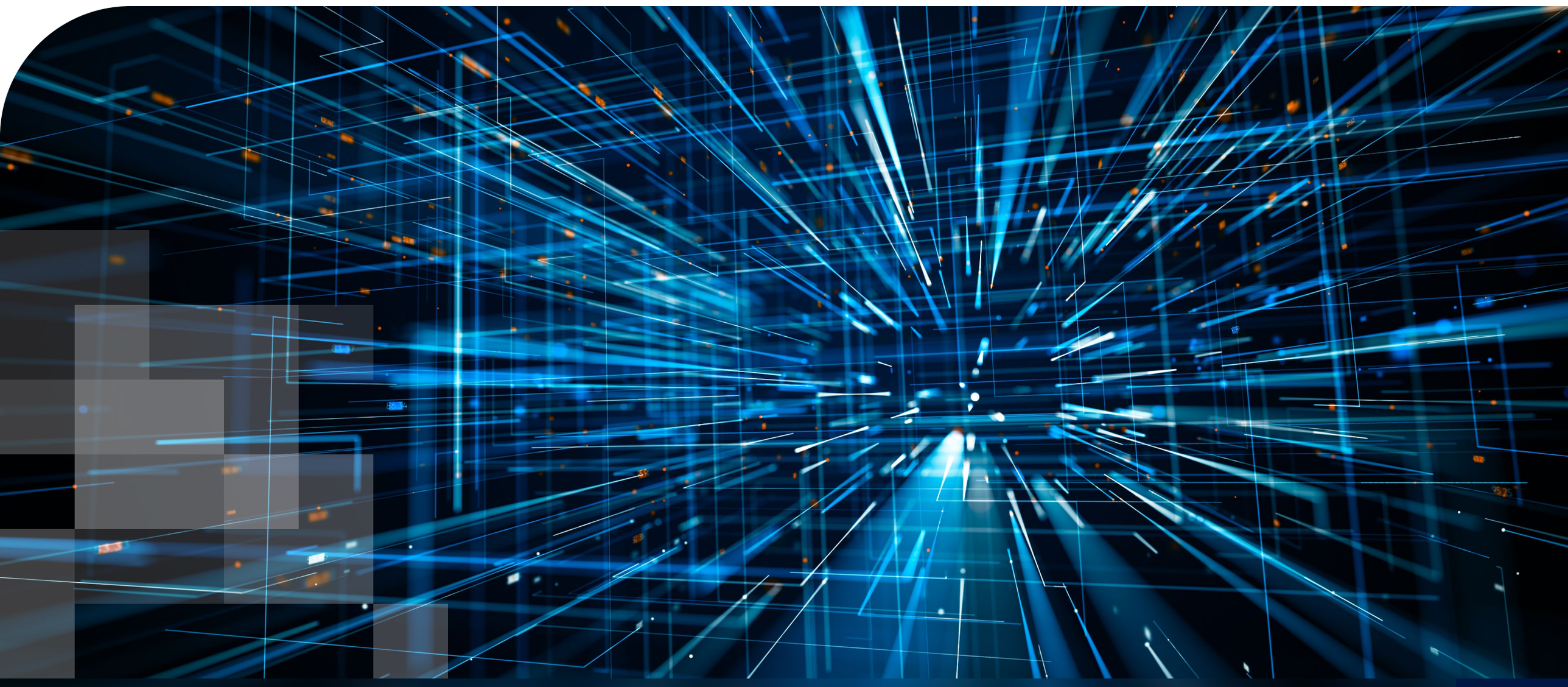
You cannot protect what you cannot see. Identifying sensitive data systems is necessary because:

Defining the Blast Radius: Security teams need to know which systems, if compromised, would result in a material breach.

Regulatory Compliance: Frameworks like GDPR, HIPAA, and PCI-DSS require organizations to maintain an inventory of where regulated data resides. Without this, audits will fail.

Prioritizing Controls: Resources are finite. Identification allows you to apply the most stringent “Zero Trust” policies to high-value assets (e.g., the HR database) while allowing more flexibility for low-risk systems.

Reducing “Data Sprawl”: Organizations often store redundant, obsolete, or trivial (ROT) data. Identifying these systems allows for data minimization—deleting what you don’t need to reduce the attack surface.



Common Difficulties

The “Identification” phase is often the most time-consuming part of a data security project due to:

Shadow IT & SaaS Sprawl: Employees often use unsanctioned tools (e.g., a personal Dropbox or a free PDF converter) to handle corporate data, moving it outside the view of IT.

Data at Rest vs. Data in Motion: Many tools can see data being uploaded (in motion) but have no visibility into the terabytes of “dark data” sitting in forgotten S3 buckets or old file shares (at rest).

Encrypted Blind Spots: Over 90% of web traffic is encrypted. If you aren’t inspecting SSL/TLS traffic, you cannot identify sensitive data moving to external systems.

Unstructured Data: It is easy to find a “Credit Card Number” in a SQL database field. It is much harder to find sensitive intellectual property hidden in a 50-page PDF or a CAD drawing.

EXAMPLES

SCENARIO	THE IDENTIFICATION CHALLENGE	THE RISK
THE “GHOST” S3 BUCKET	A developer creates a public AWS S3 bucket for a quick test and populates it with a “sample” of production customer data.	The data is unencrypted and publicly accessible via a simple URL.
PII IN COLLABORATION TOOLS	An HR employee copies a spreadsheet of employee IDs and salaries into a Slack channel or a Jira ticket to resolve a payroll issue.	This data is now indexed in a third-party SaaS platform, often without DLP monitoring.
LEGACY FILE SHARES	A “Marketing” folder on an internal NAS contains project files from 2018 that happen to include unmasked customer email lists.	An attacker moving laterally through the network finds this “low-hanging fruit.”

How Zscaler Supports Identification

Zscaler provides a “holistic” identification layer by looking at data from multiple vantage points:

ZSCALER DSPM (DATA SECURITY POSTURE MANAGEMENT)

Role: Identifies data **at rest** in public clouds (AWS, Azure, GCP).

Function: It automatically crawls your cloud environment to find “Shadow Data”—databases or buckets that IT didn’t know existed—and uses AI to classify the sensitivity of the data inside them.

ZSCALER CASB (CLOUD ACCESS SECURITY BROKER)

Role: Identifies data in **SaaS Applications** (M365, Salesforce, Box, etc.).

Function: It scans sanctioned SaaS tenants to find sensitive files that are shared publicly or with external users, providing an “Inventory” of sensitive SaaS assets.

ZSCALER DLP (DATA LOSS PREVENTION) — INLINE

Role: Identifies data **in motion.**

Function: By performing full SSL Inspection, Zscaler sees data as it moves toward the internet. It can identify that a user is uploading “Sensitive Project Code” to a personal GitHub, even if that system was previously unknown to IT.

ZSCALER PRIVATE ACCESS (ZPA) + APP DISCOVERY

Role: Identifies **Internal Systems.**

Function: ZPA can “discover” private applications running in your data center or VPC. By observing traffic patterns, it helps you identify which internal servers are actually being used to host sensitive applications, allowing you to move them from a flat network into a segmented Zero Trust environment.

SHADOW IT DISCOVERY

Role: Identifies **Unsanctioned Systems.**

Function: Zscaler logs show every single URL and cloud service your users access. It provides a “Shadow IT Report” that ranks apps by risk, helping you identify systems like “Mega.nz” or “WeTransfer” where sensitive data might be leaking.

Active & Passive Technical Discovery

CASB & SSPM Analysis: Use a Cloud Access Security Broker (CASB) or SaaS Security Posture Management (SSPM) App Governance tool to crawl your tenant environments (O365, Salesforce, ServiceNow). This identifies "Shadow IT" where users may have provisioned their own instances.

Log Aggregation (SIEM/EDR): Analyze egress logs to see where data is being sent. If 50GB of data moves to an unknown IP daily, that endpoint is now a "system of interest."

Infrastructure-as-Code (IaC) Review: For cloud-native organizations, scanning Terraform or CloudFormation templates reveals the architecture of databases and storage buckets before they are even deployed.

Administrative & Process Discovery

The Procurement Trail: Review accounts payable records for the last 24 months. Any vendor categorized under "Software," "Analytics," or "Storage" represents a potential system containing sensitive data.

Business Impact Analysis (BIA): Leverage existing Disaster Recovery documentation. BIA documents usually list the "critical" systems needed to keep the business running; these almost always contain sensitive data.

Network & API Mapping

Example: Running an API Discovery tool on your internal gateway may reveal an undocumented "Legacy_Customer_API" that still pulls from an old SQL database containing clear-text PII.

How to find APIs: Zscaler Internet Access (ZIA) and Zscaler Private Access (ZPA) act as inline sensors that convert raw network telemetry into a structured map of your API ecosystem. By analyzing traffic at the application layer (Layer 7), they identify programmatic interfaces that traditional network tools miss.

CORE DISCOVERY VIA LOG ANALYSIS

URL Pattern Matching:
Zscaler inspects the Request URI to find common API structural markers.

- **Path Indicators:** Filtering for strings such as `/api/`, `/rest/`, `/graphql/`, or `/v1/`, `/v2/`, and `/v3/`.
- **Versioning:** Identifying legacy "Zombie APIs" by spotting older versioning strings (e.g., a server still accepting `/v1/` traffic when the official standard is `/v4/`).

Content-Type Header Inspection:
Zscaler analyzes the "Content-Type" and "Accept" headers in the HTTP metadata.

- **JSON & XML:** Flagging any transaction where the header is set to `application/json`, `application/xml`, or `text/xml`.
- **Binary/Structured Formats:** Identifying more specialized API traffic like `application/x-protobuf` or `application/grpc`.

ADVANCED DISCOVERY INDICATORS (“OTHER IDEAS”)

Headless User-Agent Analysis:

- Standard web traffic identifies as a browser (Chrome, Safari).
- API discovery flags “Headless” clients such as `PostmanRuntime`, `python-requests`, `curl`, `Go-http-client`, or `Java/1.8.0`. This is the most reliable way to distinguish a human browsing a site from a script calling an API.

HTTP Method Distribution:

- While humans mostly “GET” data, APIs frequently use “POST,” “PUT,” “PATCH,” and “DELETE” to modify data.
- High volumes of these “write” methods to a specific endpoint are a primary indicator of an API backend.

Error Response Analysis:

- Monitoring for specific API error codes like `401 Unauthorized` (failed API key) or `429 Too Many Requests` (API rate limiting). These responses often reveal the presence of an API gateway or protective layer that IT may not have documented.

Part II: Defining and Verifying System Ownership

Ownership is often the most contentious part of data governance. We must distinguish between **Accountability** and **Responsibility.**

Why is this Important? (The “Accountability Gap”)

In many organizations, IT is the **custodian** (they keep the servers running), but the Business Unit is the **owner** (they understand the value and risk of the data).

Contextual Policy Making: Only a system owner can decide if a specific data flow is a “critical business process” or a “security violation.”

Access Governance: Owners are responsible for “User Access Reviews” (UAR). They must certify that the 50 people with access to the Finance folder still actually need it.

Incident Response: When a breach occurs, the “Mean Time to Respond” (MTTR) is dictated by how fast you can find the person who knows what that specific database contains and how it impacts operations.

Regulatory Requirement: GDPR (Article 30) and other frameworks require a “Record of Processing Activities” (ROPA), which must include the identity of the data controller/owner.

What Matters in this Step?

Distinguishing Roles: You must differentiate between the Technical Owner (the SysAdmin who manages the VM) and the Business Owner (the VP of Sales who owns the CRM data).

The “Orphaned System” Problem: Systems often lose owners when employees leave or departments merge. Identification of “orphaned” data is a major priority here.

Verification (Attestation): It is not enough to have a name in a database. Ownership must be verified annually. If the owner doesn’t acknowledge the system, it should be considered a candidate for decommissioning.

Why Customers Should Pay Attention

Avoiding “Security Friction”: Security teams often get blamed for “breaking the business.” By defining owners, security becomes a partner. The owner signs off on the risk, and the security team simply enforces the owner’s requirements.

Operational Efficiency: When a system is identified as having “no owner” and “no recent traffic,” it can be shut down, saving significant licensing and infrastructure costs.

Audit Readiness: Auditors don’t just ask “is the data encrypted?”; they ask “who authorized this user’s access?” If you can’t answer the second question, you fail.

Practical Tooling for Ownership & Governance

Organizations typically use these specialized tools to track and verify ownership:

ITSM & CMDB: ServiceNow or Atlassian Jira Service Management. These are the primary “Systems of Record” where the owner’s name is officially stored.

GRC Platforms (Governance, Risk, and Compliance): OneTrust, RSA Archer, or MetricStream. These tools automate the “attestation” process (sending an email to an owner asking, “Do you still own this?”).

Identity Governance (IGA): SailPoint or Okta Identity Governance. These tools link “Systems” to “Users” and “Owners” to manage the lifecycle of access.

How Zscaler Supports Ownership Verification

While Zscaler is a security platform, it provides the “ground truth” telemetry needed to verify if your ownership records match reality.

ZSCALER PRIVATE ACCESS (ZPA) + USER ACTIVITY

The Support: ZPA can show you exactly who is accessing a specific private application.

The Value: If your CMDB says the “Marketing Dept” owns a server, but ZPA logs show only “DevOps” users are accessing it, Zscaler has identified an Ownership Misalignment. This allows the customer to correct their records.

DSPM (DATA SECURITY POSTURE MANAGEMENT)

The Support: DSPM automatically pulls metadata from AWS, Azure, and GCP.

The Value: It identifies the “Creator” of a resource (e.g., an S3 bucket) and maps it to their IAM identity. It can highlight “Orphaned Data”—sensitive storage buckets created by users who are no longer with the company.

ZSCALER CASB (TENANT GOVERNANCE)

The Support: CASB identifies every SaaS application in use and who the administrators are.

The Value: It helps identify “Shadow IT Owners.” If an employee started a corporate “Miro” account with their own credit card, Zscaler identifies that user as the de facto owner of that system, allowing IT to bring it under official governance.

ZSCALER DIGITAL EXPERIENCE (ZDX)

The Support: ZDX monitors the end-to-end performance of applications.

The Value: By looking at which departments are experiencing performance issues with a specific app, you can infer which business units are the primary stakeholders (owners) of that service.

The Dual-Ownership Model

To ensure clarity, every system in your matrix must have two distinct owners:

ROLE TYPE	TITLE	RESPONSIBILITY
BUSINESS OWNER (THE STEWARD)	VP of Sales, Head of HR, Controller	**Accountability:** They own the risk. They decide who should have access and how long data is retained based on business needs.
TECHNICAL OWNER (THE CUSTODIAN)	IT Manager, DevOps Lead, DBA	**Responsibility:** They own the implementation. They manage patching, backups, encryption, and the actual granting/revoking of access.

Verification Workflows

The "Orphaned System" Test: If a system has no clear owner, IT should (following a warning period) restrict access or "scream-test" the system. The person who calls the helpdesk to complain is usually the Business Owner.

Annual Certification: Owners must "re-certify" their systems annually, confirming they are still the owner and that the system still serves a valid business purpose.

Part III: Identifying Specific Sensitive Data Elements

Knowing that a system is sensitive is not enough; you must know why it is sensitive.

Why is this Important? (The "Context" Problem)

Without classification, security policies are "blind." You cannot write a rule that says "Block sensitive data" because the computer doesn't know what "sensitive" looks like until you define it.

Precision Protection: You want to allow a user to upload a "Marketing Brochure" to LinkedIn but block them from uploading a "Customer Price List." Discovery/Classification is the only way to tell the difference.

Regulatory Granularity: GDPR requires different handling for "General PII" vs. "Sensitive Personal Data" (like health or religious info). You must classify to comply.

Cost Optimization: Not all data needs the same level of backup, encryption, or residency. Classifying data as "Public," "Internal," or "Confidential" allows you to spend your security budget where it matters most.

What Matters in this Step?

Automation at Scale: Manual tagging (asking users to label every file) is a guaranteed failure. People forget, or they label everything "Public" to avoid hassle. Automated Discovery is mandatory.

False Positive Reduction: If your discovery tool triggers an alert every time it sees a string of 16 numbers (which might be a credit card or just a part serial number), the security team will eventually ignore all alerts.

Contextual Intelligence: Knowing what the data is is only half the battle. You need to know the context: Who created it? Where is it going? Is it being moved by a human or a bot?

Why Customers Should Pay Attention

The "Shadow Data" Risk: 80% of corporate data is unstructured (Word docs, PDFs, screenshots). This data often sits in places it shouldn't be. Discovery finds the "hidden" risks that a simple system inventory (Part I) misses.

Zero Trust Maturity: A true Zero Trust architecture requires Data-Centric Security. You can't have a Zero Trust "Data" pillar if you haven't classified the data.

AI Readiness: As companies adopt GenAI (like ChatGPT or Copilot), the risk of "Data Leakage via Prompt" sky-rockets. You must classify your data before you allow employees to interact with AI, so the AI can be restricted from ingesting "Secret" data.

How Zscaler Supports Data Discovery & Classification

Zscaler acts as the **central nervous system** for data intelligence because it sees data in every state: at rest, in motion, and in use.

ZSCALER UNIFIED DLP ENGINE

The Support: Zscaler uses a single engine to classify data across Endpoint, Web, SaaS, and Cloud.

The Value: You define your "Sensitive Data Profile" (e.g., "German PII") once, and it is automatically discovered across all channels. This eliminates the need to manage five different DLP tools.

AI-POWERED DATA DISCOVERY

The Support: Zscaler uses Machine Learning (ML) to automatically categorize data into 200+ categories (e.g., Financial documents, Legal contracts, Medical records).

The Value: Customers don't have to write complex "Regular Expressions" (RegEx). Zscaler's AI "understands" what a resume looks like versus a source code file, drastically reducing false positives.

EXACT DATA MATCHING (EDM) & FINGERPRINTING (IDM)

The Support: Zscaler can "fingerprint" a customer's specific database (e.g., their actual customer list) or a specific document template (e.g., a proprietary design spec).

The Value: If a user tries to exfiltrate even a small snippet of an actual customer record, Zscaler recognizes it with 100% accuracy because it matches the "fingerprint" of the original data.

ZSCALER DSPM (DATA SECURITY POSTURE MANAGEMENT)

The Support: DSPM scans "at-rest" data in AWS S3, Azure Blobs, and Google Cloud Storage.

The Value: It discovers sensitive data that is "sitting still." It tells the customer: "You have 5,000 credit card numbers in an unencrypted S3 bucket that hasn't been accessed in 2 years."

OPTICAL CHARACTER RECOGNITION (OCR)

The Support: Zscaler can "read" text inside images, such as photos of passports, credit cards, or screenshots of sensitive spreadsheets.

The Value: This closes a massive gap where employees try to bypass DLP by taking a screenshot of sensitive data instead of copying the text.

Content Inspection Techniques

Pattern Matching (Regex): Identifying structured data like Credit Card Numbers (Luhn check) or Social Security Numbers.

Exact Data Matching (EDM): You provide a "fingerprint" of your actual customer database to your discovery tool. The tool then finds those specific records (e.g., "Company Confidential") elsewhere, reducing false positives.

Optical Character Recognition (OCR): Essential for scanning "image-based" sensitive data, such as scanned passports in a HR folder or handwritten checks in a finance share.

Contextual Analysis

Proximity Analysis: Finding a 16-digit number is a "hit." Finding a 16-digit number near the words "EXP DATE" and "CVV" is a "critical incident."

File Metadata: Sensitive data often hides in the metadata of "Public" documents (e.g., the "Comments" or "Author" fields of a PDF).

Part IV: The Data/System Matrix (The Master Record)

This matrix is the final deliverable. It should be a dynamic document (integrated into a GRC tool or a managed Database) rather than a static PDF.

The Comprehensive Matrix Template

SYSTEM ID	SYSTEM NAME	HOSTING TYPE	BUSINESS OWNER	TECHNICAL OWNER	PRIMARY DATA TYPES	CLASSIFICATION	REGULATORY SCOPE	ENCRYPTION STATUS
SYS-001	Workday	SaaS (Public)	Jane Doe (VP HR)	Bill Smith (IT Mgr)	PII, Salary, Bank Info	**Restricted**	GDPR, CCPA	At Rest / In Transit
SYS-002	Customer DB	AWS (Private)	John Roe (VP Sales)	Amy Lin (DevOps)	PII, PCI, Usage Logs	**Confidential**	PCI-DSS, GDPR	At Rest
SYS-003	Engineering Repo	On-Prem	Mike S. (Dir Arch)	Tech Lead	Source Code, IP	**Restricted**	IP Protection	In Transit Only
SYS-004	Marketing Blog	SaaS (Public)	CMO	Webmaster	Public Articles	**Public**	None	None

How to Maintain the Matrix

1 INGESTION

New systems are added during the “Vendor Risk Assessment” phase of procurement.

2 VALIDATION

Quarterly automated scans verify if the “Primary Data Types” have changed (e.g., if a system that previously held only names starts holding Credit Card data).

3 DEPRECATION

When a system is decommissioned, the “Data Disposal Certificate” must be linked to the Matrix entry before the record is archived.

Part V: Designing High-Fidelity Detections for Sensitive Data

Creating effective data loss prevention (DLP) detections is a balancing act between high sensitivity (catching every leak) and high precision (minimizing false positives). A successful detection strategy moves from simple keyword matching to advanced, context-aware technologies that understand the "intent" of data movement.

Data Analysis: What to Detect?

Before writing a single rule, customers must analyze their "Crown Jewels" to determine their digital DNA.

Regulated Data (PII, PCI, PHI): This includes structured patterns like Social Security Numbers, Credit Card Numbers (PAN), and IBANs. Most of these follow international standards (e.g., the Luhn algorithm for credit cards).

Intellectual Property (IP): This is often unstructured. It includes source code, CAD drawings, M&A strategy documents, and proprietary recipes or formulas.

Contextual Metadata: Sometimes the sensitivity is not in the content, but the context—such as files originating from a "Highly Confidential" folder or documents containing headers like "Internal Use Only."

Building the Foundation: Keywords and Regular Expressions (RegEx)

Keywords and RegEx are the “first responders” of data detection.

Keyword Dictionaries:
Customers should compile lists of high-value terms. For example, a legal firm might use “Attorney-Client Privilege,” while a tech company might use project code names (e.g., “Project Titan”). find common API structural markers.

Regular Expressions (RegEx):
These are used for data that follows a predictable pattern but has variable values.

- Example: To find a custom employee ID that always starts with two letters, a dash, and four numbers, a customer would write: `[A-Z]{2}-\d{4}.`

Proximity Analysis:
To reduce false positives, Zscaler allows customers to use “Proximity” rules. Instead of just looking for a 16-digit number, the rule only triggers if the word “Visa” or “Credit Card” is found within 10 words of that number.

Advanced Zscaler Detection Technologies

When keywords and RegEx are not enough—especially for unstructured or massive datasets—Zscaler provides “Next-Gen” detection engines.

EXACT DATA MATCHING (EDM)

How it Works: A customer exports their actual database (e.g., a CSV of 1 million customer records) and uploads a “fingerprinted” hash to Zscaler.

The Value: Zscaler can now detect the exfiltration of a specific customer’s record. If a user tries to upload a file containing “John Doe” and his specific birthdate and account number, Zscaler triggers a match with 100% accuracy.

INDEXED DOCUMENT MATCHING (IDM)

How it Works: Customers provide “template” documents, such as a blank patent application or a standard financial audit form.

The Value: Even if a user changes 20% of the text in a document, IDM recognizes the “DNA” of the original template and blocks the upload. This is ideal for protecting proprietary forms and intellectual property.

AI & MACHINE LEARNING (ML) CLASSIFICATION

How it Works: Zscaler uses pre-trained ML models to recognize categories of data without any manual configuration.

The Value: The engine can automatically identify “Finance Documents,” “Resumes,” or “Legal Contracts” simply by analyzing the structure and linguistics of the file. This protects against “Shadow Data” that the security team didn’t even know existed. These are great for an initial discovery on potentially sensitive data in an identified asset but should not be used alone in enforcing data leakage prevention. Another identifier needs to be added, e.g. Finance Document (ML) and IBAN (predefined dictionary).

OPTICAL CHARACTER RECOGNITION (OCR)

How it Works: The Zscaler engine “reads” the text inside images and screenshots.

The Value: This prevents “Data Leaks via Camera” or screenshots. If an employee takes a screenshot of a sensitive database and tries to send it via personal webmail, Zscaler extracts the text from the image and applies the standard DLP rules.

Conclusion

The journey through these three phases—Identification, Ownership, and Classification—represents a fundamental shift from traditional, perimeter-based security to a modern, data-centric Zero Trust posture. By moving through this methodology, an organization transforms its data security from a series of reactive, fragmented fire drills into a proactive and automated governance engine. It begins with the critical "Visibility Phase," where the goal is to map the entire digital landscape and illuminate the "dark data" or Shadow IT that often sits outside the view of traditional IT controls. Without this initial clarity, any security policy is merely guesswork, leaving the organization vulnerable to blind spots in cloud storage, unsanctioned SaaS apps, and encrypted traffic.

Once the landscape is mapped, the focus shifts to "Accountability." Defining and verifying system ownership bridges the gap between technical infrastructure and business reality. This phase ensures that every data repository has a designated stakeholder who understands its value and can make informed decisions about risk and access. By establishing clear lines of responsibility, the security team stops being a bottleneck and instead becomes an enabler, enforcing the business-driven requirements of the actual data owners. This structural alignment is what allows a security program to scale effectively without creating unnecessary friction for the end-user or the business unit.

The final layer of intelligence is found in "Discovery and Classification," where the actual content of the data is analyzed and tagged. This is where the "what" meets the "how," allowing for surgical precision in security policies. By leveraging AI and machine learning to distinguish between public information and highly sensitive intellectual property, organizations can apply the most stringent protections exactly where they are needed most. This granular intelligence is especially vital in an era of rapid AI adoption, where the risk of sensitive data leaking into LLMs or being exfiltrated via screenshots is higher than ever before.

Zscaler's unique value in this entire process lies in its ability to unify these three disparate steps into a single, cohesive platform. Rather than managing a complex web of point products for discovery, governance, and enforcement, Zscaler provides a "single pane of glass" that sees data at rest in the cloud, in motion across the web, and in use within private applications. For a leader like yourself, the message to the customer is clear: Zscaler doesn't just provide a tool; it provides a comprehensive Data Security Platform that automates the journey from total "Data Blindness" to a state of "Content-Aware Zero Trust." This approach ultimately reduces the attack surface, simplifies compliance, and ensures that the most valuable corporate assets remain secure, no matter where they travel.

About Zscaler

Zscaler (NASDAQ: ZS) is a pioneer and global leader in zero trust security. The world's largest businesses, critical infrastructure organizations, and government agencies rely on Zscaler to secure users, branches, applications, data & devices, and to accelerate digital transformation initiatives. Distributed across more than 160 data centers globally, the Zscaler Zero Trust Exchange™ platform combined with advanced AI combats billions of cyber threats and policy violations every day and unlocks productivity gains for modern enterprises by reducing costs and complexity.

+1 408.533.0288 **Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134** **zscaler.com**

zscaler™

Act Fast.
Stay Secure.