ASSESSMENT WORKBOOK

UNEXPECTED DATA EXPOSURE CHECKLIST

A practical checklist to identify, assess and reduce your unexpected data risk



GROUND LABS

Know where to look.

Most organizations know where sensitive data ***should*** live: CRMs, payment environments and HR systems with clear ownership, access controls and retention policies.

Where it ***actually*** ends up is less clear. Customers share data through unintended channels, staff paste sensitive information into everyday tools, and systems create copies in logs, backups and archives. Over time, these copies accumulate.

WHAT IS UNEXPECTED DATA?

Sensitive information held in locations an organization doesn't know about and doesn't monitor, without the robust controls of its core platforms.

In this guide, we uncover the ten locations unexpected data most commonly accumulates, how it gets there and what you can do to manage it.

HOW TO USE THIS WORKBOOK

01 **Scan the checklist**
Mark the review status of each of the 10 locations.

02 **Investigate each location**
Use the practical prompts as a starting point for discovery.

03 **Record your findings**
Capture evidence, a responsible owner and a clear next action.

04 **Prioritize follow-up**
Move unresolved items to the action plan at the end.

Where might sensitive data hide?

For each location, select one status. Use the location pages that follow to guide the review. “Reviewed” indicates a completed check, not that the location is risk-free.

LOCATION	REVIEWED	INVESTIGATE	N/A
01 Customer inbound channels	☐	☐	☐
02 Collaboration tools & ticketing platforms	☐	☐	☐
03 Source code & CI/CD pipelines	☐	☐	☐
04 Internal email	☐	☐	☐
05 Logs and diagnostic data	☐	☐	☐
06 File shares & document libraries	☐	☐	☐
07 AI prompts & chat histories	☐	☐	☐
08 User endpoints & sync folders	☐	☐	☐
09 Backups & archives	☐	☐	☐
10 Legacy & decommissioned systems	☐	☐	☐

Next: use the detailed pages to find what to examine, then log evidence and your next action.

01 Customer inbound channels

WHY THIS LOCATION MATTERS

Customers don't always use the channels designed for them. They email photos of their payment cards, attach passport scans and medical letters to complaints and claims, type account details into chat windows and read card numbers aloud on recorded calls.

All of this lands in support systems, shared inboxes, call recordings and the file stores behind them, then gets copied into customer records, reporting systems and AI tools.

This data is often unsolicited and in many cases actively discouraged. As a result, it falls outside governance processes and accumulates over time. Under GDPR and similar laws, businesses are responsible for this data whether they requested it or not.

WHERE TO LOOK

Support inboxes, chat transcripts, call recordings and uploaded attachments.

YOUR INVESTIGATION CHECKLIST

- [ ] **Check where unsolicited sensitive information lands.**
- [ ] **Identify downstream copies and who can access them.**
- [ ] **Review handling and retention procedures for inbound content.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

02 Collaboration tools and ticketing platforms

WHY THIS LOCATION MATTERS

Modern businesses are high-pressure environments, where individuals are measured on performance and output. With chat tools and ticketing systems on hand to help solve problems, staff will provide whatever information they need to get the job done as quickly and efficiently as possible in a relatively unfiltered way. After all, these tools are internal.

Ticketing systems and collaboration tools are heavily integrated into operational workflows, connected to systems and third-party applications with wide access to the sensitive data within them.

WHERE TO LOOK

Team chat, service desks, issue tickets and connected applications.

YOUR INVESTIGATION CHECKLIST

- [ ] **Look for pasted customer information, credentials and extracts.**
- [ ] **Review application integrations and access permissions.**
- [ ] **Check whether sensitive content is retained beyond its purpose.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

03 Source code and CI/CD pipelines

WHY THIS LOCATION MATTERS

Source code and CI/CD tools are common locations for stored secrets, bypassing separate secrets manager services for convenience and simplicity. Unfortunately, where these end up in released code they can be used to gain unauthorized access to or exposure of data in connected systems.

Even where these are removed from release code, log trails for development code still capture them, along with the associated configuration files, packages software images and legacy code versions.

WHERE TO LOOK

Repositories, pipeline definitions, build logs and configuration files.

YOUR INVESTIGATION CHECKLIST

- [ ] **Check for credentials, tokens and other embedded secrets.**
- [ ] **Include historical commits, build artifacts and logs in scope.**
- [ ] **Confirm a process exists to remove and rotate exposed secrets.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

04 Internal email

WHY THIS LOCATION MATTERS

It's not just inbound email that leads to unexpected stores of sensitive data. While organizations may implement inbound filtering to block specific types of data, these rules don't typically apply internally.

Staff copy sensitive information into emails and include system exports or reports as attachments, sharing them internally or forwarding on to suppliers. These records soon pile up as user mailboxes are rarely cleared with automated retention policies.

WHERE TO LOOK

Employee mailboxes, shared inboxes, forwarded messages and attachments.

YOUR INVESTIGATION CHECKLIST

- [ ] **Review common export and attachment-sharing patterns.**
- [ ] **Check internal and supplier forwarding pathways.**
- [ ] **Assess mailbox retention and access controls.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

05 Logs and diagnostic data

WHY THIS LOCATION MATTERS

System logs are crucial to support troubleshooting and security investigations. However, there's the potential for vast amounts of sensitive information to be reported in log records and retained unknowingly for extended periods.

Even automated snapshot files saved by software when it crashes can preserve sensitive information hidden in directories on its host.

In well-monitored environments, system logs are aggregated into centralized SIEM solutions and kept for long periods in the event of an incident or event where they may be needed for investigation.

WHERE TO LOOK

Application logs, diagnostic snapshots, SIEM repositories.

YOUR INVESTIGATION CHECKLIST

- [ ] **Check whether logs record sensitive fields or payloads.**
- [ ] **Inspect error reports and automated crash artifacts.**
- [ ] **Review log access and retention periods.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

06 File shares and document libraries

WHY THIS LOCATION MATTERS

Shared drives and document libraries are where most organizations keep their documents, whether on in-house file servers or in cloud platforms such as SharePoint, Teams, OneDrive and Google Drive.

Files can be shared with anyone who has the link. These stores typically hold personnel records, scanned ID documents, financial data, intellectual property and deal material.

WHERE TO LOOK

Shared drives, SharePoint, Teams, OneDrive and Google Drive.

YOUR INVESTIGATION CHECKLIST

- [ ] **Find sensitive files in broadly accessible folders.**
- [ ] **Review link-based sharing and external access.**
- [ ] **Identify outdated copies and retention gaps.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

07 AI prompts and chat histories

WHY THIS LOCATION MATTERS

Employees are increasingly encouraged to use AI tools for greater productivity and lower staffing needs. As with ticketing and support queries, this drives staff to share code, contracts and customer records with AI tools.

Prompt and chat histories build up over time, capturing extensive amounts of sensitive information. Many of these AI tools can be connected by users to other corporate systems, expanding the amount of sensitive information accessible to them.

WHERE TO LOOK

AI chat histories, saved prompts and connected data sources.

YOUR INVESTIGATION CHECKLIST

- [ ] **Determine which AI tools and integrations are in use.**
- [ ] **Check for sensitive content in prompt and chat histories.**
- [ ] **Review access, retention and approved-use guidance.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

08 User endpoints and sync folders

WHY THIS LOCATION MATTERS

For many people, almost everything they do at work is via a workstation or laptop. They become mini storage units of our workday, complete with local versions of the corporate cloud systems we rely on to function.

End point security tools focus on verifying the configuration of the device and stopping malware, but remain agnostic to the accumulation of sensitive information hosted on the machine.

Many more staff now work between home and office, capturing local copies of company information to ensure they can be effective without needing to use a slow VPN or while dealing with a flaky internet connection.

WHERE TO LOOK

Laptops, desktops, download folders and synced directories.

YOUR INVESTIGATION CHECKLIST

- [ ] **Check local caches, exported reports and synced copies.**
- [ ] **Assess visibility into sensitive files on endpoints.**
- [ ] **Review cleanup and offboarding procedures.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

09 Backups and archives

WHY THIS LOCATION MATTERS

Backups copy everything in the systems they protect, including unexpected data. Archives hold old email, documents and database extracts kept "just in case."

Backups and archives are managed as recovery and compliance tools rather than as data stores, so deletion requests and automated enforcement policies rarely reach them.

WHERE TO LOOK

Backup sets, email archives and historical data exports.

YOUR INVESTIGATION CHECKLIST

- [ ] **Review sensitive data contained in backup scope.**
- [ ] **Check archival retention, access and restoration rules.**
- [ ] **Consider how deletion requirements apply to copies.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

10 Legacy and decommissioned systems

WHY THIS LOCATION MATTERS

Businesses can't always let go of old systems, even when their replacements are up and running. Instead the old platform is retained for historical lookups, or the data from it archived in case it's needed, regardless of the sensitivity of the information it held.

Over time, they become invisible, lost and forgotten parts of the digital estate, decommissioned but never completely removed.

WHERE TO LOOK

Retired applications, historical databases and migration archives.

YOUR INVESTIGATION CHECKLIST

- [ ] **Identify legacy platforms and ownership gaps.**
- [ ] **Review what sensitive data remains accessible.**
- [ ] **Document a retention, migration or disposal decision.**

DATE: _____

EVIDENCE / NEXT ACTION: _____

TAKE CONTROL OF UNEXPECTED DATA WITH ENTERPRISE RECON

Managing unexpected data requires **data intelligence:** an evidence-based understanding of what sensitive data exists, where it sits, who can access it and what exposure it represents to the business.

Enterprise Recon from Ground Labs, builds that intelligence by finding these unexpected data and identifying the data within them, across structured and unstructured data in on-premises and cloud environments.

HOW ENTERPRISE RECON SUPPORTS DISCOVERY

- **Find unexpected sensitive data across structured and unstructured sources.**
- **Identify sensitive content across on-premises and cloud environments.**
- **Use 300+ preconfigured data types and define custom types with GLASS Studio.**

GREATER VISIBILITY | SIMPLIFIED COMPLIANCE | LOWER RISK

Learn more: groundlabs.com/enterprise-recon