Genesys DATA

Service Description

Automated Data Discovery

Scope

Automated Data Discovery (ADD) is the process of using AI, pattern recognition, and metadata analysis to automatically locate, classify, and map data across both structured and unstructured repositories in cloud and on premises environments. It identifies information that exists but remains unmanaged, unclassified, or forgotten, such as legacy files or unsecured collaboration data containing personally identifiable information, sensitive content, or overexposed business records.

The process extends across a wide range of repositories including file shares, cloud collaboration platforms, email systems, databases, archives, CRM systems, and SaaS applications. It covers diverse data types such as personal, financial, credential, intellectual property, and regulatory data, ensuring comprehensive visibility across both active and dormant datasets in production and sandbox environments.

ADD directly supports core security principles of confidentiality, integrity, and accountability (CIA). It strengthens confidentiality by identifying where sensitive data such as Personally Identifiable Information (PII), Payment Card Industry (PCI) data, or Protected Health Information (PHI) resides and determining who has access, enabling immediate remediation of public or overly broad exposure. It enhances integrity by detecting data copies, outdated or mislabelled files, and inconsistent access controls that could undermine data reliability. Finally, it reinforces accountability through detailed audit trails and logging of access, modification, and sharing activities, ensuring ownership, traceability, and responsible data handling.

An additional benefit of the service, ADD also reveals data that is redundant, obsolete, or trivial (ROT). Such data often accumulates unnoticed across repositories, consuming storage, increasing operational costs, and elevating security and compliance risks. With growing emphasis on data minimisation and lifecycle management, the ability to identify and remediate ROT is increasingly important to ensure compliance with retention and disposal requirements. Proactively addressing ROT improves governance efficiency, reduces risk exposure, and supports a more streamlined, less costly and defensible data environment.

Approach

This process systematically identifies target systems and data sensitivities, automates scanning and classification to detect sensitive patterns, maps exposure and access

risks, correlates findings with behavioural insights, and validates results with stakeholders to prioritise remediation and establish ongoing monitoring:

- **Planning & Scoping:** Identify target systems, data sensitivity tiers, data size (scan and visualise) and regulatory frameworks (GDPR, APRA CPS 234, ISO 27001).
- **Automated Scanning & Classification:** Use rule-based and AI-driven classification to detect sensitive patterns (names, card numbers, secrets, PHI).
- **Exposure Mapping:** Identify public links, guest access, excessive permissions, and data shared externally.
- **Risk Correlation:** Combine discovery results with behaviour analytics (e.g., abnormal data access) for context.
- **Validation & Reporting:** Confirm findings with stakeholders, prioritise remediation, and establish continuous monitoring.

Discovery & Risk Mapping

As part of performing this service - a secure SaaS tenancy of the Data Governance / DSPM (Data Security Posture Management) platform will be provisioned to enable data discovery, classification, signalling, and file recognition across in-scope data sets.

Connectivity will be established through tenant onboarding, delegated app consent, and Entra ID role-based access providing controlled and auditable integration with the client environment.

The DSPM is IRAP-assessed and SOC 2 Type II certified, ensuring alignment with recognised information-security and assurance frameworks. All traffic is encrypted in transit and governed under the client's enterprise security, privacy, and compliance controls.

Value

Automated Data Discovery delivers visibility into hidden data, strengthens security, ensures compliance, improves efficiency, and embeds lifecycle governance for a resilient data environment. Protects sensitive data, customer trust, and intellectual property from accidental or negligent use:

- **Illuminate Hidden Risk:** Uncovers and classifies dark and unstructured data, giving full visibility into where sensitive information resides and who can access it.
- **Strengthen Data Security:** Reduces exposure by identifying and remediating open links, excessive permissions, and stale accounts across cloud and on-prem systems.
- **Enable Regulatory Confidence:** Automates classification, retention, and audit evidence generation, demonstrating compliance with APRA, GDPR, and ISO standards.
- **Drive Operational Efficiency:** Automates data discovery, tagging, and remediation workflows, cutting manual effort and improving response times by up to 70%.

- **Establish Lifecycle Governance:** Embeds retention, immutability, and metadata tagging to deliver a 90% improvement in overall data-posture visibility after the first assessment.

Benefits

These are the key advantages of undertaking this service offering:

- Improves visibility and control over dark and unclassified data across all repositories.
- Reduces unnecessary exposure and access risks by tightening permissions and removing public links.
- Enhances regulatory and audit readiness through automated classification, retention, and reporting.
- Increases operational efficiency by streamlining remediation and policy enforcement.
- Optimises storage and cost efficiency through the removal of redundant, obsolete, and trivial data.
- Strengthens lifecycle management by ensuring data is properly tagged, retained, and disposed of in line with governance requirements.

Outputs

These outputs are delivered to support implementation and forward planning:

- **Data Discovery and Classification Report**
  - Comprehensive inventory of structured and unstructured data across cloud, on premises, and legacy repositories – limited to in-scope data sets.
  - Automated classification of sensitive data (PII, PHI, PCI, credentials, intellectual property) using AI and rule based scanning.
  - Visual dashboards detailing data volumes, sensitivity levels, ownership, and geographic location.
- **Data Exposure and Access Risk Register**
  - Catalogue of exposed or at-risk data, including public, guest, organisation wide, or cross border access.
  - Identification of over privileged identities and access inconsistencies.
- **Executive Summary and Visual Dashboard**
  - Board level summary highlighting key findings, measurable improvements, and next step priorities.
  - Real time dashboards visualising data risk posture, remediation progress, and compliance mapping.
  - Benchmark scorecard showing visibility improvements and exposure reduction achieved through automation.

Genesys Data offers the following additional outputs in conjunction with a DSPM deployment (purchased as a subscription offering).

- **OPTIONAL: Remediation Plan and DSPM Integration Roadmap**
  - Structured plan outlining remediation actions and dependencies.
  - Integration roadmap aligning findings with a DSPM platform.
- **OPTIONAL: Data Lifecycle and Retention Model**
  - Framework for end-to-end data lifecycle management, from creation and classification to lawful retention and disposal.
  - Integration of immutability controls (e.g. WORM, version locking) and metadata tagging to maintain policy consistency.
  - Retention rules mapped to applicable regulatory and organisational obligations, ensuring defensible governance.
- **OPTIONAL: Governance and Compliance Alignment**
  - Mapping of data categories and risks to relevant regulatory frameworks such as APRA CPS 234/235, GDPR, ISO 27001, and NIST CSF.
  - Identification of compliance gaps in retention, immutability, and metadata tagging.
- **OPTIONAL:** Prioritised remediation plan with assigned ownership, actions, and resolution timeframes.

Capabilities & Constraints

In Scope Data-Sets

In scope data-sets comprise all structured and unstructured information repositories, systems, and storage locations included for discovery under this engagement. These data sets encompass approved sources such as file shares, collaboration platforms, cloud storage, email systems, archives, databases, and SaaS applications identified during planning and onboarding. Only these defined environments are subject to automated discovery, classification, and risk assessment within the secure tenancy.

Data Volume Coverage (Scan and Classification)

Data volume coverage refers to the total quantity of data, measured in terabytes (TB), scheduled to be scanned during the Automated Data Discovery process, and the defined subset of that volume selected for classification and visualisation. The subset represents data requiring deeper content inspection, including optical character recognition (OCR) and semantic data recognition, extending beyond metadata analysis to identify sensitive or high value information for governance and reporting purposes.

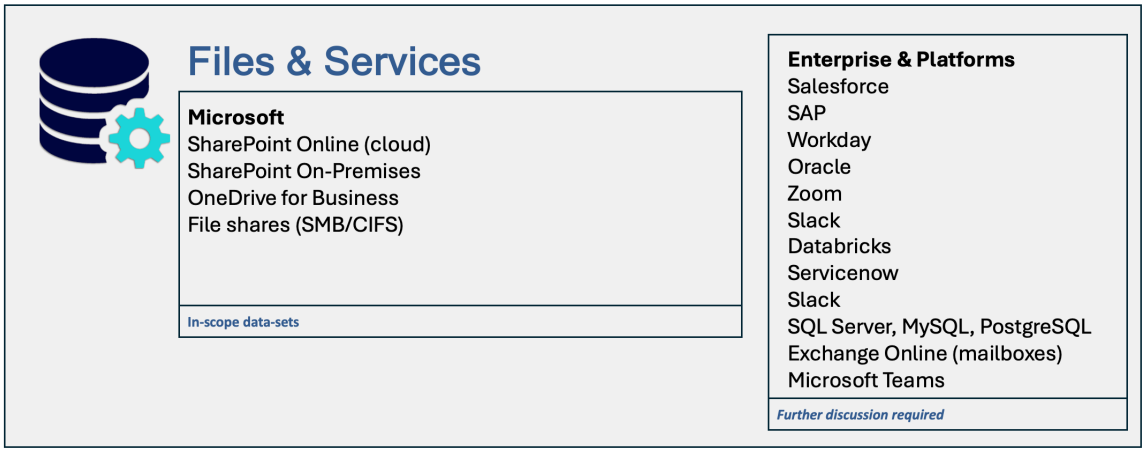

Figure 1 - represents the in-scope Data Sets; estimated data volumes (subject to scanning, classification, and visualisation) are to be agreed.

Genesys Data stands with you to transform how to measure, approach and harness the power of your data, offering a portfolio of services that address the complexities of - getting data right! By integrating cutting-edge AI data-discovery techniques, advancing maturity modelling, and offering Data, Privacy and Security by Design, you'll be able to demonstrate appropriate duty-of-care, exceed privacy obligations, effectively gain control of data and understand how to become more cyber resilient and AI ready!

Learn more at www.genesysdata.com.au

Copyright © 2025 Genesys Data. All rights reserved. Other names may be trademarks of their respective owners.