Here's the plain text version of the Spirion SDP Solution Overview document:

Spirion SDP Solution Overview

Prepared by: Rob Server

February 14, 2025

Table of Contents

Executive Summary

Spirion is honored to be considered for <PROSPECT>'s data discovery and classification initiative.

In 2006, Todd Feinman and David Goldman, started Spirion with an audacious goal to eliminate date breaches and the pain associated with them. Today, that guiding principal remains our paramount focus – to protect what matters most – the sensitive personal and private data of an organization's customers, colleagues, partners, and communities. For more than 15 years, Spirion has solved sensitive data discovery, classification, and remediation challenges for thousands of global customers, proving scalability and performance from mid-sized to the largest of enterprises.

Spirion's methodology has always been to attack data at the source— at rest. Moreover, Gartner and Forrester, like Spirion, recognize that the cornerstone to a sound data protection program begins with accurate data discovery to locate ALL sensitive data as opposed to just PII. Spirion provides an industry leading 98.5% accuracy rate of data discovery right out-of-the-box thanks to our proprietary AnyFind™ technology Tolly Test Report, 2021. Reducing the number of false positives is imperative in enabling organizations to take the appropriate protective measures on their data. Therefore, accuracy, automation, and actionability are Spirion's major market differentiators which help organizations meet data requirements of various frameworks, regulations, and standards such as NIST, CMMC, GDPR, FERPA, HIPAA, and PCI-DSS.

Automation and ease of use are critical factors to consider when deploying discovery and classification technology, as both factors accelerate time to value and strengthen adoption of data protection programs. Spirion excels in helping enterprises develop classification schemas (if needed), leveraging our platform to automate and operationalize them. Business needs should guide the defining of classifications while Spirion's solution imposes relevant policies via automated workflows that act on files once classified, where desired. Spirion's accuracy, coupled with automation, eliminates uncertainty, and will mitigate <PROSPECT>'s overall risk profile as it relates to sensitive data security and privacy.

At Spirion, we want every individual to treat sensitive data protection seriously, because every lost or stolen record containing sensitive data can cause real harm to the individuals/organizations whose data is compromised. While the vendor landscape provides a diverse and often complex, convoluted set of offerings for data security and data privacy solutions, we urge organizations to focus on Step 1 – Sensitive Data Discovery of ALL sensitive data. You cannot protect what you do not know exists.

We look forward to taking the next steps with <PROSPECT> on this journey to protect what matters most.

Sincerely, <Name>

Privacy-Grade™ Data Protection

If you want to protect every type of sensitive data, it is crucial to find it wherever it resides. This capability defines Privacy-Grade™, which ensures you can accurately discover the sensitive data entrusted to your organization.

While typical security solutions focus on the perimeter, reacting to intrusion and data breaches at the network level, Privacy-Grade solutions from Spirion take a data centric, proactive approach. They address data at rest, protecting it at the root—down to the individual data record level—thereby enhancing the efficacy of perimeter-focused solutions.

Spirion's Sensitive Data Platform is unmatched in its ability to find more types of data in more places than any other solution on the planet. And by continually innovating, Spirion empowers organizations to protect what matters most, ensuring comprehensive data security from the inside out.

Raising the data privacy and security bar to address today's realities

Globally, the trend towards stronger data privacy regulations continues, with numerous countries updating or introducing new laws to protect consumer data.

As of 2023, over 120 countries have enacted data privacy laws, reflecting a global commitment to protecting personal data. In the U.S. alone, eight states passed privacy legislation in 2023, and five of those laws go into effect in 2024.

Data continues to expand exponentially, and breach volumes have grown right alongside it.

In 2023, more than 353 million individuals had their sensitive data compromised.

The average cost of a data breach rose to $4.88 million in 2023, marking a 10% increase and the highest cost since the COVID pandemic.

3 Steps to Preemptively Protect Sensitive Data

Detecting, containing, and remediating data breaches takes even longer when companies don't know where their data is stored, who has access to it, and where their weak attack vectors are located. All regulations, frameworks, and standards associated with data security require understanding what data there is first, then apply controls. If compliance to a framework, regulation, or standard such as NIST, CMMC, GDPR, FERPA, HIPAA, or PCI-DSS is needed, the first step is to identify and understand where all sensitive data lives and what it is.

Step 1 - Locate ALL Sensitive Data:

Without end-to-end visibility across all endpoints and systems within your IT environment, there may be unknown data at risk of a breach. New records are created every second, and data moves through systems faster than any person can track. Understanding where data lives is a critical first step to reducing exposure. Once endpoints with access to high value data are

known, teams can easily recognize and strengthen attack vectors to prevent unauthorized access.

Step 2 - Classify and Catalog Sensitive Data:

Discovering data records is valuable, but only with an understanding of what information that data contains. When 80%-90% of a company's data is unstructured, many organizations don't know what data they're retaining, making remediating a breach and compliance reporting much more challenging.

Classifying and cataloging the sensitive data across all systems is critical to maintaining a strong security posture. By tagging records for specific collection, storage, access, and security parameters, organizations can more effectively manage their growing data footprint. When using and protecting sensitive records, context can make a huge difference as to how that data should be managed. Inconsistent data management often leads to mis-categorization, causing further confusion when a breach occurs. The persistent application of meaningful, context-rich classification labels is critical once areas of risk are discovered.

Step 3 - Remediate Unnecessarily Exposed Sensitive Data:

After data is located and cataloged, companies should take control of where and how these sensitive records are stored. Creating security controls and limiting where information is located can reduce the risk of unnecessary exposure.

Identifying where data is hiding provides extensive benefits to IT teams. For example, if data shows up somewhere it doesn't belong, that can indicate where administrators need to remove access or restrict systems to prevent data leaks and breaches. If data is discovered where it's not intended to be stored, that reveals an opportunity for IT to fix a weak attack vector. With an end-to-end view of what data was accessible in the event of a breach, accurate incidents and results can be reported to executive leadership, relevant government agencies, and people whose data was impacted in a timely and compliant manner.

How Spirion Can Help

Data discovery and classification projects aren't easy. We know, we've done thousands of them. The dynamic nature of data and the inherent stewardship challenges often prevent organizations from even getting started. Finding the right partner, establishing clear roles, delivering impact early on and thinking program vs. project dramatically improves the odds of ongoing success.

Data privacy is critically important in today's technology-first world, but privacy is impossible without security. Data discovery, classification, and remediation can be arduous and time-consuming processes; Spirion saves time and resources by protecting sensitive data automatically.

With Spirion, organizations can discover the complete landscape of sensitive structured and unstructured data across their IT infrastructure, including on networks, in the cloud, on remote file servers, and on physical devices, with industry-leading 98.5% accuracy. The contextual, automated discovery process captures company data wherever it resides, eliminating blind spots to reduce the risk of breach or unauthorized access without interrupting day-to-day business operations.

Once data is discovered, automated data classification allows analysts to better understand their data without applying compliance and security rules manually. Alongside reducing the risk of human error, Spirion can incorporate automated logic to embed persistent labels that interoperate with in-motion data controls while also operationalizing policies for remediation, protection, and user awareness throughout the data lifecycle.

Spirion is Step One for reducing your exposure to a potential breach. Spirion gives you clarity as to what sensitive data you have and where it is located, control over how your data is stored and used, and confidence that your data is protected. It begins with our proven 98.5% accurate discovery and then enforced through our powerful and purposeful automated classification and remediation capabilities.

Together with our Technology Alliance Partners, Spirion provides easy, immediate, and automated protection of the sensitive information that is "low-hanging fruit" for attackers while improving regulatory compliance, lowering organizational risk profile and liability, and enhancing consumer brand confidence.

Our Methodology

Spirion's methodology has always been to attack data at the source, at rest. Moreover, Gartner and Forrester, like Spirion, recognize that the cornerstone to a sound data protection program begins with accurate data discovery (Step 1).

Automation and ease of use are critical factors to consider when deploying discovery and classification technology, as both factors accelerate time to value and strengthen adoption of data protection programs. Spirion excels in helping enterprises develop classification schemas (if needed), leveraging our platform to automate and operationalize them. Business needs should guide the defining of classifications while Spirion's solution imposes relevant policies via automated workflows that act on files once classified, where desired.

Solution Overview: Sensitive Data Platform

Spirion Sensitive Data Platform (SDP) provides Privacy-Grade™ data discovery and purposeful classification in a highly scalable SaaS hybrid architecture, able to thoroughly scan both on-premises endpoints/servers and cloud repositories at enterprise scale. SDP quickly and automatically discovers, classifies, and remediates almost any form of sensitive data or personally identifiable information (PII) – including custom data types – anywhere it resides.

Privacy-Grade Discovery: Identifying Sensitive Data with AnyFind™ and Sensitive Data Engine®

Featuring proprietary Context-Aware Data Interrogation Algorithms (CADIA™), Spirion's AnyFind™ technology identifies, remotely and locally, both structured and unstructured sensitive data across cloud storage, laptops, file servers, SaaS applications, and throughout organization's files, emails, databases, websites, Microsoft SharePoint sites, and more, all with the industry's highest precision.

Spirion identifies and classifies hard-to-find sensitive data, such as personally identifiable information, nonpublic data, intellectual property, and toxic identifiers that lead to data breaches, reputational damage, and financial loss.

Hyper-accurate, customizable, pre-trained, and pre-configured CADIA models identify: Social Security numbers and worldwide national identifiers, credit card data, bank account and other financial information, driver licenses and state IDs, passport numbers, passwords, dates of birth and other healthcare-related data.

The Key is the CADIA Model

In partnership with academia since 2006, Spirion's CADIA algorithms have been developed and trained on exabytes of highly confidential, diverse, and high-quality data.

Given today's data-centric regulatory environment, this kind of real-world training cannot be replicated.

CADIA produces highly accurate and repeatable results through procedural validation. By learning from your data to automatically extend the search parameters, CADIA leverages a decision tree branching process to determine future results.

Those results are further correlated through additional content and context to ensure validity.

CADIA's model leverages diverse data and location-specific techniques including predefined and linear classifiers, adaptive contextual awareness, proximity-based content evaluation, procedural validators, multi-tiered decision tree branching, checksums, decision trees, exact data matches, dictionaries, and additional validation techniques.

The result: Spirion avoids the high false-positive rate, expense, and frustration associated with pure pattern matching and regular expressions (RegEx) frequently seen in data leakage prevention tools.

What Is CADIA?

CADIA was purpose-built to identify sensitive data with accurate and repeatable results, avoiding many of the costly issues associated with general-purpose AI and ML-based

applications. CADIA employs "human-in-the-loop" (HITL) machine learning, where humans ensure that the model's predictions are accurate and finely tuned to maintain the highest level of precision and repeatability.

Finding sensitive information using purely AI technologies requires large volumes of high-quality, diverse, sensitive data sets, but it can still lead to responses that may not be consistent from one analysis to another. By leveraging CADIA for sensitive data discovery and classification, Spirion delivers repeatable true-positive identifications with an extremely high degree of accuracy and the lowest false-negative rate.

Obtaining Accuracy in CADIA

CADIA applies numerous validation algorithms before presenting results to an end user. As one algorithm finishes, Spirion determines which validator to run next.

With an SSN, for example, you might see a high-group check to determine if the SSN could have been issued by the U.S. Social Security Administration before June 25, 2011 (13-year-olds and older). Or, if the SSN does not have dashes, Spirion might determine that the SSN is unformatted and look for context to determine if it is truly an SSN in a database column, an SSN in a sentence, or a 9-digit number that is not an SSN.

CADIA continues to apply the correct validation algorithms until it is sure it has found a true positive or eliminated a false positive. CADIA will change its behavior and search criteria based on previous results automatically, assuming those features are enabled.

As another example, if you are searching for passwords on an endpoint and find some in web browsers, Spirion takes those results and uses them to search for new results in other locations, such as emails or files.

Tunable Classifiers for Custom Data Types

In addition, Spirion Sensitive Data Engine enables you to discover any intellectual property and other private data unique to your organization, such as user and device IDs, as well as identifiers associated with mobile applications and other custom software.

With highly customizable context options, you can build proprietary sensitive datatypes with exceptional accuracy. Sensitive Data Engine has an application programming interface (API) that allows developers to create their own rules and definitions for finding personal information and sensitive data.

This extensible system allows for uniquely defined logic that enables complex evaluation systems and algorithms.

Context-Rich Sensitive Data Classification

Purposeful: Today, organizations recognize that comprehensive data privacy and security isn't just about identifying or discovering sensitive information; it's also a matter of properly and purposefully classifying data to protect it against unauthorized access, use, and modification that can lead to severe financial, regulatory, and legal ramifications.

Automated: Spirion automatically labels data based on its purpose of collection, the process through which it was collected, and its privacy level. The labels are then federated across the entire IT environment where enforcement activities can be taken based upon company policies.

Persistent: Traditional classification methods only look for data in motion, which means that existing data that's been sitting around goes unprotected. Spirion analyzes and classifies data at rest, and those classifications follow the data as it's in motion. Spirion-applied labels can prevent email technology from sending it, or users from copying it for potentially malicious purposes, by interoperating with in-motion data controls (such as NGFWs, CASBs, DLPs).

Contextual: Six out-of-the-box persistent, contextual classification categories provide flexibility in how they can organize and define their data to stay compliant, including: sensitivity, purposeful data collection, business processes, applicable regulatory guidelines, consumer preferences, and custom categories.

Interoperable: Spirion classifications are imbedded in files as easily readable metadata. With Spirion AnyFind™ accuracy and playbook automation, this context-rich metadata can easily enhance third-party downstream systems such as traditional DLP, Firewalls, SSE, xDR, DRM, DAG, SIEM solutions, etc.

Spirion and Generative AI

Spirion ensures that your AI and LLM data retains value without becoming dangerously exposed. Leverage our world-leading Privacy-Grade™ accuracy in identifying your sensitive data and combine it with our unique context-rich classification engine to label your data as LLM-ready. Spirion can then supercharge your in-motion data loss prevention (DLP) by using GenAI LLM-ready data tags available in the Spirion console, can be applied manually or through the automation of Spirion's data privacy playbooks, ensuring actionable classification that builds AI Governance to protect your AI initiatives.

Spirion classifications include a globally unique identifier (GUID), which can then be consumed by your DLP to ensure the proper protective measures are enacted based on the sensitivity of a file as determined by a Spirion scan. Then, only data that is considered LLM-ready is utilized in the training or end-user implementation of AI tools.

Data Privacy Playbooks

Automated: SDP can be configured to operate in a nearly fully automated fashion using Data Privacy Playbooks. Preconfigured search polices can be scheduled to search for any target, at any time, on a recurring basis, returning data back to the console while scans are being

performed. Search results trigger (optionally) automatic, logically defined outcomes through classification and remediation actions.

Dynamic: Diverse sets of actions can be configured depending on a wide variety of factors, from file location, data content, or ACL attributes, to when files were last modified or accessed. Actions are defined via interactive cards that provide all relevant options while hiding everything irrelevant to the user's task at that moment.

Intuitive: Playbooks operationalize information security policies using progressive disclosure by adding single cards, one at a time in the Playbook builder, and use self-segmentation in the Playbook viewer to empower the user to intuitively choose the appropriate path from discovery through remediation.

Sensitive Data Remediation

Actionable: SDP delivers privacy-grade remediation solutions that enable a business to securely process sensitive information. SDP's data processing actions include, but are not limited to, the collection, retention, logging, generation, transformation, use, disclosure, sharing, and disposal of personal data.

Automated: SDP can automatically remediate any information with unparalleled granularity to address varying levels of data sensitivity and use cases. Users can proactively identify and tag data with predefined scopes, for various regulations, earmarking it for the appropriate level of remediation at any time, no matter where the data resides within an organization. Additionally, custom remediation can be executed via SDP's "Execute Script" function. SDP can conduct passive (report only) remediation and active remediation (direct action on files, for example).

Persistent Classification

Spirion can apply persistent metadata tags to supported files automatically via Playbook. This can inform downstream reporting and edge systems about files that are moving around and possibly getting replicated to other systems. It also provides a visual indicator to end users to inform them that a given file has been classified.

Quarantine

Spirion has a quarantine feature that will by Playbook automatically move a file to another location. It securely deletes the original file and leaves a stub file with a message to the file owner.

Redact

Spirion can, via Playbook, automatically redact data out of Microsoft Office documents and text files. This can remove sensitive matches and leave the rest of the document intact.

Restrict Access

Spirion is not a DAG solution, however we can remove excessive permissions automatically via Playbooks. In the scenario where a workstation user does not have local administrator access, it is viable for administrators to Restrict Access – "locking" a file in place. This optimally will limit access to the SYSTEM account and the Administrator.

Apply Microsoft Purview Labels

With Spirion's integration with Microsoft Purview, common customers can utilize Spirion Accuracy and Playbook Automation to apply Purview labels where Microsoft cannot – directly on the desktop and other on-prem resources! Purview labeling can provide significant functionality – even at the E3/A3 license level, Rights Managed encryption is possible, preventing data from being accessed by unauthorized users.

Ignore

Allows users to mark false positives (e.g., a non-sensitive number mistaken for a Social Security number) so that Spirion skips these items in future scans. This is not a remediation action per se but helps refine the process by reducing unnecessary alerts.

Execute Script

Spirion Playbooks support integration with custom PowerShell or Python scripts, allowing organizations to execute bespoke remediation actions tailored to their needs. For example, a script could move files to a third-party secure storage solution or integrate with an external workflow system. There are many examples of Scripts in the Spirion Marketplace https://www.spirion.com/marketplace

Reporting & Executive Dashboards

SPIglass™ Executive Dashboard: Sensitive Private Information (SPIglass™) Dashboard presents sensitive data risk in financial terms that are meaningful to board members, executive leadership, and other stakeholders (Is it a $200,000 risk or a $25 million risk?). SPIglass empowers data-driven decision-making, demonstrating progress, and highlighting pockets of risk in the security posture.

SDV³ Dashboard: Sensitive Data Risk (SDV³) Dashboard provides a quantitative measure of data risk that is directly tied to the sensitivity of personal data stored on IT systems. Spotlighting the riskiest data assets, so organizations can objectively manage trade-offs and quantify risk according to the value of the date found, the volume of that data, and the vulnerability of the asset the data resides on.

Spirion Custom Reporting and Restful API: SDP offers out-of-the-box reports, dashboards, and custom reporting. Additionally, Spirion has a Swagger restful API great for programmatically

exporting data for analysis in external data modeling tools such as PowerBI and enriching other rest enabled SIEM systems.

Integrations & Interoperability

Once data is classified, the SDP platform has the capability to continuously locate specified data through persistent classification. This foundational capability, combined with the ability to integrate into other security tools, optimizes risk mitigation.

SDP seamlessly interoperates with other security tools such as traditional DLPs, Information/Data Rights Management solutions, Security Information Event Management systems, De-identification/Synthetic Data tools, and Rights Management solutions. SDP has several native API integrations; however, any solution that can perform content inspection can easily read Spirion's embedded classification metadata.

Spirion partners with many leading security providers to deliver integrated solutions that reduce complexity and help achieve data protection goals faster and more effectively.

Scalability & Performance

SDP allows for local-node, remote, or a hybrid model for deployments. On-prem nodes are deployed to workstations, PCs, or other local computer platforms. They make highly effective use of already available computer resources and the high bandwidth/low contention storage buses connecting disk drives to the server or PC they are running on. Only the scan results are returned to the SDP Console, which greatly reduces network bandwidth and storage issues.

SDP Nodes: The SDP node is typically installed on physical and/or virtual endpoints (laptops & desktops) and other machines (servers) designated as Discovery Team members. The node is responsible for performing the actual scan. The console controls most node activity. However, the node GUI does provide functionality at the endpoint level for users to participate in sensitive data protection activities (if so desired).

Distributed Scanning: To address the sensitive data management challenges faced by contemporary data-driven enterprises, Spirion recently has overhauled its sensitive data scanning engine to speed up the scanning of repositories and returning results. The initial phase, released in February, introduces a multi-threaded architecture capable of processing scans two to five times faster than before. Future phases will further amplify scanning speeds, reaching hundreds of times faster.

Discovery teams are fully automated and highly scalable. By simply selecting a set of nodes to use, SDP divides and distributes the scan process across the discovery team, sharing a global scan history to ensure that no work is duplicated. As nodes become available, they check a queue for the next available scan portion and begin scanning immediately.

There are several advantages to using SDP Distributed Scanning to scan assets:

- Improved Distributed Scanning: Uses multiple agents to scan a single large-volume target, ensuring faster scans by efficiently dividing the locations to be scanned and eliminating time consuming "discovery" phase.
- Results Streaming: Processed results are streamed in real-time, enabling immediate actions on sensitive data.
- Status / Health Reporting: Users can monitor the progress and health of their scans in real-time.
- Enhanced Encryption: This includes generating separate passwords for each search, with only the designated agent being able to read it.

Spirion's Proven Delivery Process

At Spirion, consumer data protection and customer success are at the heart of what we do. Our core competencies focus on enabling organizations to protect their intellectual property and the ever-growing amount of sensitive data housed across their infrastructure. We strive to ensure optimal product usage with both data security and privacy stakeholders, to deliver the best return on investment, through our Proven Process.

Customer Support

Quick Start Customer Onboarding

Every Spirion customer is assigned a dedicated Professional Services Engineer (PSE) for the 12-week Quick Start Customer Onboarding process (4-weeks of Foundations, followed by an 8-week Customization project plan). The challenges Spirion's solutions solve for are intrinsically difficult, and with over 15 years of experience, we know that at the end of the day your success is ours, making it our paramount focus to set every customer up for success on day one.

Premium Support

Spirion offers Premium Support options to assist customers beyond the Quick Start Customer Onboarding process and throughout the journey to protect what matters most. Our Premium Support customers are assigned to a dedicated Technical Account Manager (TAM), assisting with best practices and strategies for scaling usage across the information ecosystem and maturing through the programmatic approach.

Customer Success

Regardless of the industry or customer size, Spirion delivers personalized, ongoing relationship support, by assigning every customer to a dedicated Customer Success Manager (CSM). CSM's conduct Monthly Operating Reviews (MORs) with customer's technical and day-to-day operation teams, and Quarterly Business Reviews (QBRs) with customer's business stakeholders.

Technical Support

We look for meaningful ways to deliver first-class support with our online knowledge base, technical and advisory user groups, educational programs, and dedicated phone support. Priority 1 tickets are addressed with a committed one-hour response time, standard tickets are addressed within a four-hour response time, however most tickets are responded to within an average of 12 minutes.

We know our customers are not just buying a software solution; they are buying into Spirion as a company. We are a team of privacy advocates who believe in what we do and who we do it for. To us, the work we do is about protecting your colleagues, customers, and their families from a breach that could forever impact their lives. We take that very seriously and will do everything we can to help our customers protect what matters most.

Appendix

Supported Unstructured Target Locations

- Windows 10 and higher direct agent deployment support
- RHEL 7+ direct agent
- Apple macOS direct agent x64 and M1/2/3 processors
- Exchange Online and Exchange On-Prem
- SharePoint Online and SharePoint On-Prem
- OneDrive for Business
- Box
- Dropbox
- Google Workspace Gmail and Google Drive
- Any mapped drive, SMB/CIFS/NFS shares, NAS/SAN devices
- Website crawling
- Bitbucket

Supported Relational Databases

- Oracle
- MSSQL
- Sybase
- DB2
- Informix
- Interbase
- SQLBase
- SQL Anywhere
- MySQL
- SQLite
- PostgreSQL
- Mongo DB
- General OLE DB drivers

- General ODBC drivers

Supported File Types to Search for Sensitive Data

The "Search these File Types" setting allows you to let Spirion search only common file types, all filterable file types, all files, or a custom list.

[Extensive list of supported file types included in the document]

Supported AnyScan Targets

In addition to the normal targets supported by Sensitive Data Platform, a number of additional targets are supported through special drivers which establish an ODBC-like connection to the target.

[Extensive list of supported AnyScan targets included in the document]

Some Third-Party Integrations for Enhanced Remediation

Spirion SDP extends its remediation capabilities through integrations with third-party tools, allowing organizations to leverage existing security infrastructure for more advanced or specialized actions. These integrations enhance the platform's flexibility and ensure compatibility with broader data protection ecosystems:

- Encryption with Third-Party Tools (Atakama, Seclore, Thales)
- Integration with Microsoft Azure Information Protection (AIP)
- Data Loss Prevention (DLP) Systems
- Cloud Access Security Brokers (CASB)
- Information Rights Management (IRM) and Digital Rights Management (DRM)
- Script Execution (PowerShell/Python)
- De-Identification Tools