LightBeam.ai

Data Privacy Automation for Google Drive

Reference Architecture

Table of Contents

Executive Summary

The key to meeting the requirements of today's privacy regulations and protecting personal information (PI) from unauthorized use and disclosure lies in understanding and managing the use of personal information within an

organization's data environment. Spread across a multitude of repositories and application data sets, PI use can be difficult to manage through written policy alone.

We at LightBeam.ai believe the best way to implement policy across an organization is to supplement written policy with procedures for technical controls designed for specific applications and functions. By working with our clients we have developed several applications and function-specific controls focusing on discovering, analyzing, and enforcing control over the use of personal information within popular applications.

Google Drive, or Gdrive, is short for a popular data storage repository that is part of the Google suite of business products. Gdrive offers many data features, including creating, storing, processing, and sharing information within work teams. The Google suite of products like GDocs, Sheets, Slides, Forms, and more are all centrally located and designed to store files and documents in an easy-to-use information-sharing platform. In addition to ease of use, the ability to create, copy, and distribute files can also create new risks for personal information(PI) and sensitive personal information (SPI) data leakage from an organization. Similar to more traditional network share drives, the unchecked creation and duplication of files in large repositories has long been an issue with Privacy Teams trying to understand and control the use of PI in their organizations

Our AI-driven platform engine, Spectra, can easily be configured to monitor files in Gdrive and automatically discover, analyze, and enforce privacy policies regarding the use of sensitive information stored in Gdrive. By finding and either raising alerts, redacting, or deleting files that inappropriately contain PI, organizations can reduce privacy risk and meet retention requirements for data that is no longer needed. By then automating the execution of these control policies, Privacy Officers can develop custom rule sets that continually scan, monitor, and control how PI is used and controlled within Gdrive. The details of how this happens are discussed below.

Introduction

Audience

This document is intended for organizations that have implemented Google Drive and whose information processing uses personal information. It is meant for both technical and non-technical audiences. Privacy Officers, CISOs/Security Architects, and Support leaders within organizations overseeing the use of PI and Google Drive will find this reference architecture useful in automating data privacy controls.

Purpose

This document provides greater details on the problem of processing personal information within Google Drive and how LightBeam can be used to manage the use of PI and reduce the risk posed by file duplications or inappropriate and long-term storage of PI in Google Drive. Although this document is primarily about Google Drive, it applies in principle to other storage tools like MS Teams, or traditional network share drives.

Google Drive Overview

Google Drive is a file storage and synchronization service developed by Google. Launched on April 24, 2012, Google Drive allows users to store files in the cloud, synchronize files across devices, and share files among teammates. Using project teams and folders with files being created for any type of team or workgroup, Google Drive allows for communications and information sharing across resources to promote efficiency and productivity. Google Drive also integrates with other applications, like Microsoft 365 and the Google Suite, allowing for a more free flow of PI to other applications and repositories.

Often communication files stored in Google Drive can contain PI which is needed to complete transactions, process orders, or complete other activities. Used across a multitude of business types, Google Drive supports many process types. Many

customer interactions require the use of PI and the type of PI will vary by process type. While processing a transaction, PI may be required, however, long-term storage of PI in old files creates risk and should be addressed. Guided by the privacy principle to only keep PI for as long as is needed, LightBeam.ai has developed controls specific for how Google Drive uses and retains personal information.

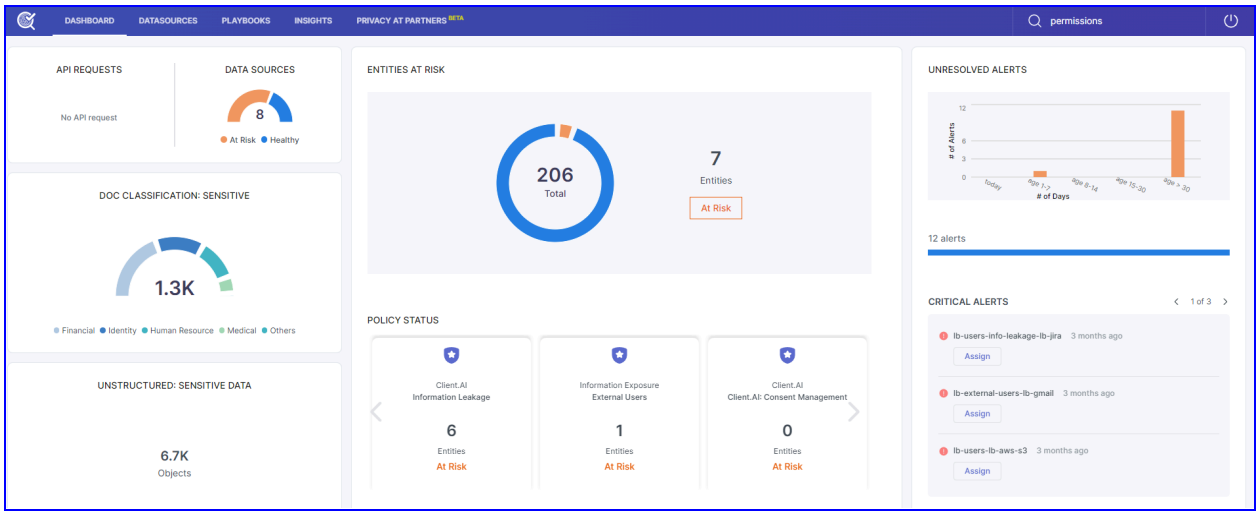
LightBeam Data Privacy Automation Platform

A pioneer in the data privacy automation (DPA) category, LightBeam is on a mission to empower organizations to access and manage their PI and SPI securely. Leveraging its identity-centric discovery & classification engine, Spectra, LightBeam ties data cataloging, access, and sharing into a unified privacy control platform.

LightBeam empowers privacy and compliance executives to keep their organizations under continuous compliance for GDPR, CCPA, HIPAA, and PCI-DSS among others, while information security executives can finally rest assured that sensitive data is being used, accessed, and stored securely.
LightBeam's 360 view of the data environment provides an up-to-date accurate dashboard of data sources, data attributes, Entities (identities), control policies, and permission lists. The following is a quick look at how LightBeam brings an unparalleled view of the PI and SPI that an organization carries within a myriad of data repositories.

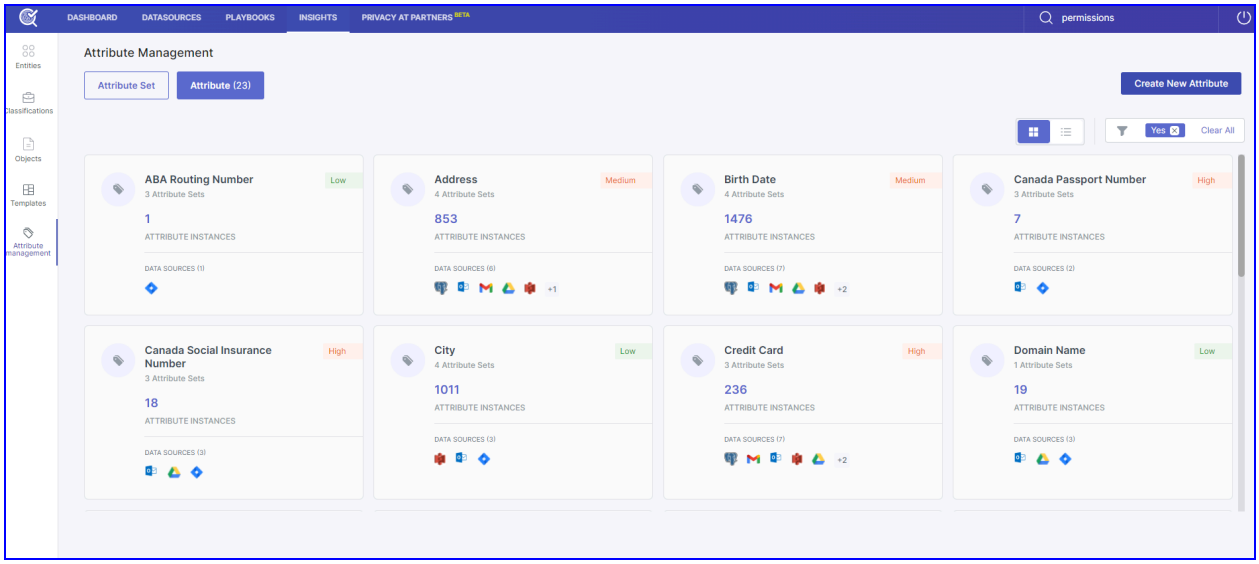
Main Dashboard

The main dashboard provides a high-level view of all data sources where sensitive data is present, the entities (customers/employees/patients et al) whose sensitive data is being carried, and any alerts that might need attention.



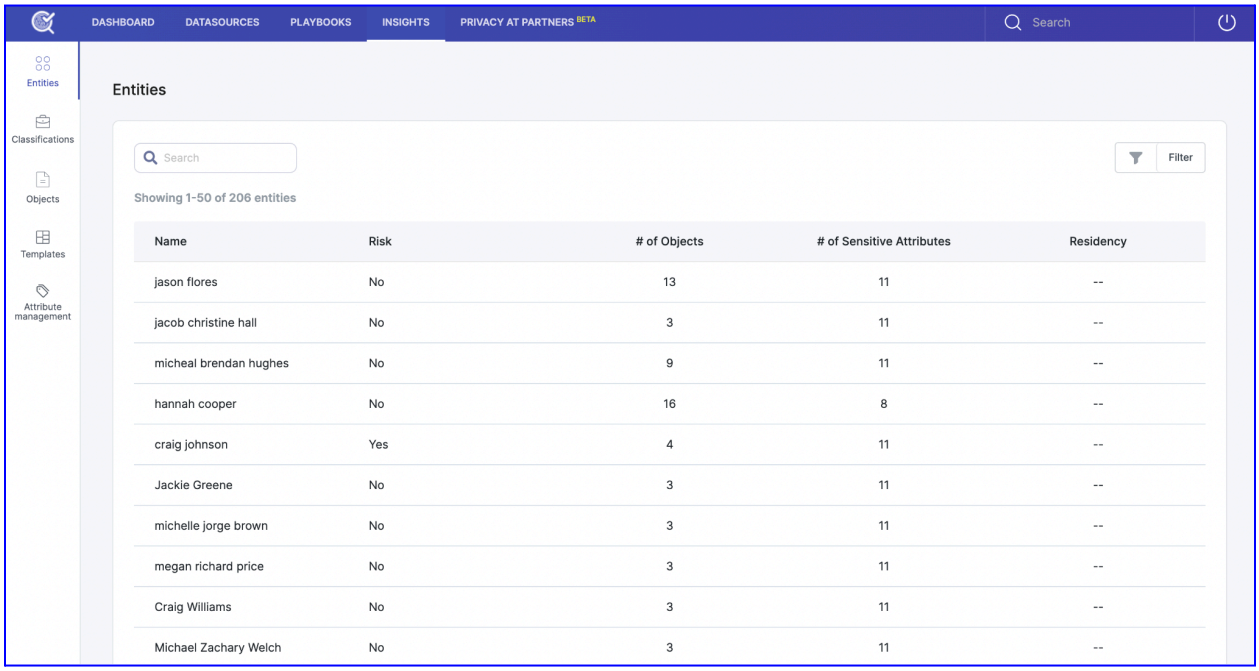
Attribute View

LightBeam has over 200 pre-configured sensitive attributes (sometimes called fields/columns) in its system and is capable of recognizing their identifiers (sometimes called Values) from all the data sources; moreover, users can also add their attributes to the system and make it learn from various sources.
These attributes have 3 sensitivity levels based on their weight in the system (i.e. high, medium & low)
Examples of attributes are U.S. Social Security Number, Loan Account Number, Medical Record Number, etc.



Entity View

Centered on the individual, the entity view provides a precise breakdown of what data is being held for any individual, in what data sources, and if there are any known associated risks. This view supports GDPR, CCPA, and other individual rights requests.



Operational Phases

LightBeam's Spectra DPA platform employs a three-phase approach to managing privacy risk. These phases include Detect, Enforce, and Automate. Each of these phases builds on the previous phase to create a fully automated privacy management system that can;

1. Understand the existence and use of PI.
2. Create control policies with resulting actions.
3. Create automated tasks to execute control policies.

Detect

LightBeam's initial step is to gain an understanding of the data environment. This includes connecting to the applications and repositories and discovering sensitive data elements called "attributes." Attributes are contained in applications and repositories and are duplicated across the environment based on the relevant business processes. LightBeam uses API connections to analyze structured and unstructured repositories and identify the data attributes, attribute types, the related sensitivity levels. Then, an Entity is resolved from the data related to an identified individual or "entity".

By understanding the data source and entity data that exists in the environment the LightBeam platform learns which data is important to an organization and its business processes. With this understanding as a foundation, LightBeam is able to then set policies as to how that data is stored, shared, and viewed.

During the detect phase, LightBeam natively recognizes and classifies;

- 200+ common attributes including the common identifiers from a variety of countries.
- Industry attribute type sets like (Financial, Healthcare, Identity…)

- Unlimited client-specific attributes - every LightBeam customer is unique and may carry sensitive data that is unique to them. LightBeam enables customers to add custom attributes. (e.g. Customer account number, employee number, member numbers, SKUs, and other values)
- Document classification of type and sensitivity of data contained in attachments in a service ticket.
- Sensitive attributes detected across multiple data repositories are linked using a machine learning algorithm to see if they belong to a single entity. The cross-linking of fragments of information to a central identity is a unique capability that helps customers understand not only if sensitive data is at risk but more importantly, whose data it is that might be at risk.
- The DETECT phase helps create data maps, RoPA reports, and a 360-degree view of all information that's present about customers within an organization's systems.
-

Enforce

The enforce phase is used to establish the rules for data usage. There are four primary control components in the enforce phase. These include Policies, Permissions, Alerts, and Redaction.

Policies

LightBeam Policies are configured to track both internal and external data sources. Each policy may contain multiple rule sets that define the search criteria and details about the data including; attribute sets and types, data sources, alert level settings, and the associated relevant regulations, and are configured via a query selection screen.
Policies include:

- Types of policies. policy types include Internal, External, and Leakage.
- The contact information for whom an alert should be sent to.
- The setting of the permissions list for whitelisting approved data sources.

White Listed data sources marked as gold source repositories are helpful in architecting a complete data schema.

Permissions

- Permission lists (also sometimes referred to as Permit Lists) establish and maintain an inventory of approved repositories and uses of PI.
    - Approved repositories are added to a permission list and used to compare new scan results.
    - Alerts can be raised when a new instance is not found on a permission list.
    - Workflows are initiated by an alert to approve new instances and update the permission list so future findings will not raise an alert.

Alerts

- Alerts are used to notify system owners and others that a policy has been violated and that action may be needed.
    - Alerts are triggered based on rule sets inside policies.
    - Alerts can be set for specific applications or all connected applications.
    - Alerts can trigger a workflow to drive a review and approve cycle

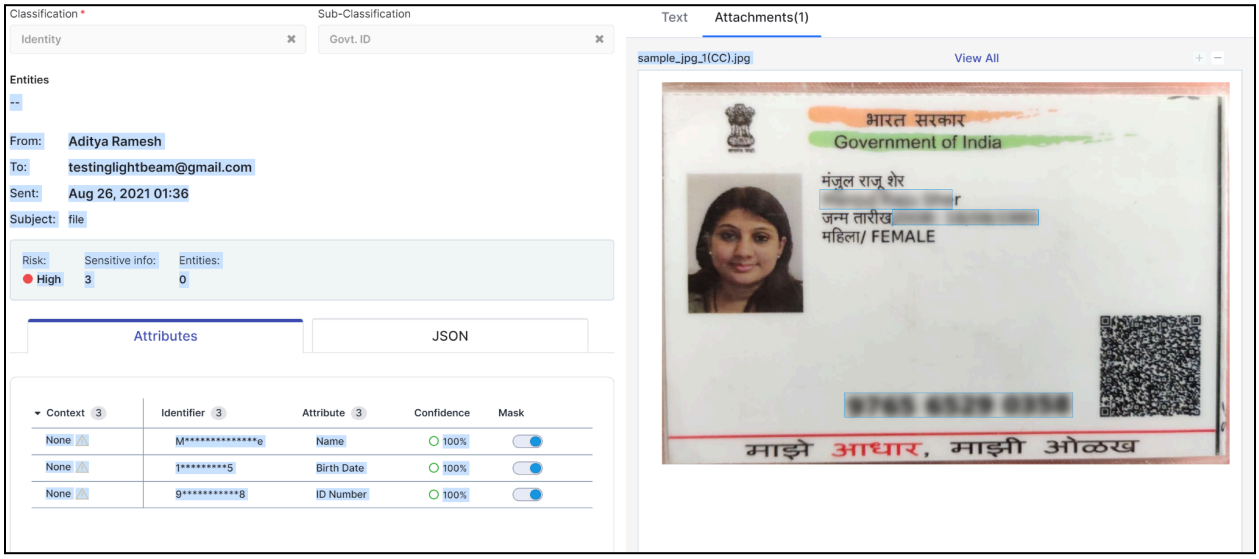
Redaction

An example of automation is LightBeam's ability to redact the presence of sensitive data within data repositories.

- The redaction control is used to maintain accurate processing records while reducing the use and risk of PI by obfuscating select fields from view.
    - Redaction control works on many document types including image files.
    - Redaction control can be set to obfuscate any sensitive data fields while keeping the rest of the process information intact in LightBeam
    - Redaction filtering can be role-based for eyes-only operations.

- The redaction control is used to maintain accurate processing records while reducing the use and risk of PI by obfuscating select fields from view.
    - Redaction control works on many document types including image files.
    - Redaction control can be set to obfuscate any sensitive data fields while keeping the rest of the process information intact in LightBeam
    - Redaction filtering can be role-based for eyes-only operations.

In the example below a government ID is redacted of sensitive fields with some less sensitive information un-redacted. LightBeam can learn and scan images as well as documents and repositories.



Guided by LightBeam's established policies, the scanning engine, Spectra, continuously scans the data environment looking for changes to the data. New copies and uses can quickly be identified and either added to a permissions list, redacted, or removed. By automating the execution of enforcement controls like alerting and redacting on a continual basis, an always-on accurate inventory of personal information is created. The process maintains an identity-centric index that can be used to facilitate the retrieval of an Individual's PI and aid in the processing of Individual Rights Requests. Additionally, DPA allows for duplicated data to be easily monitored and controlled reducing data leakage.

Data Privacy Automation for Google Drive

Automate

Utilizing LightBeams DPA technology to execute privacy process controls to continually scan, monitor, and control data usage is a powerful tool for today's Privacy, Compliance, IT, and IT Security teams. Automated monitoring of IT system controls has long been a part of most modern IT and security programs. Now the monitoring of sensitive data usage through automated processes greatly expands visibility, control, and understanding of an organization's sensitive data use across the data lifecycle.

LightBeam customers can configure policies to trigger actions to manage the use of PI stored in Google Drive. The use case is centered on Google Drive folders and the PI stored in it. Spectra can discover PI in a variety of file types including text, PDF, and image files. LightBeam will identify any of the 200+ common attributes as well as any client custom attributes stored in structured and unstructured files and databases. LightBeam will scan folders and files containing PI and based on the rule set of the control policy that is in force automatically take steps to;

1, Raise an alert and initiate a review and approve workflow

And/or

2. Redact the PI stored in the file.

As an example, A client may want to redact any sensitive information in any part of a folder after a particular time frame. To accomplish this, Spectra would first analyze a file to see if it meets a date criteria. Select all files which are past a specified date, and that also contain sensitive information. All sensitive data in the selected files

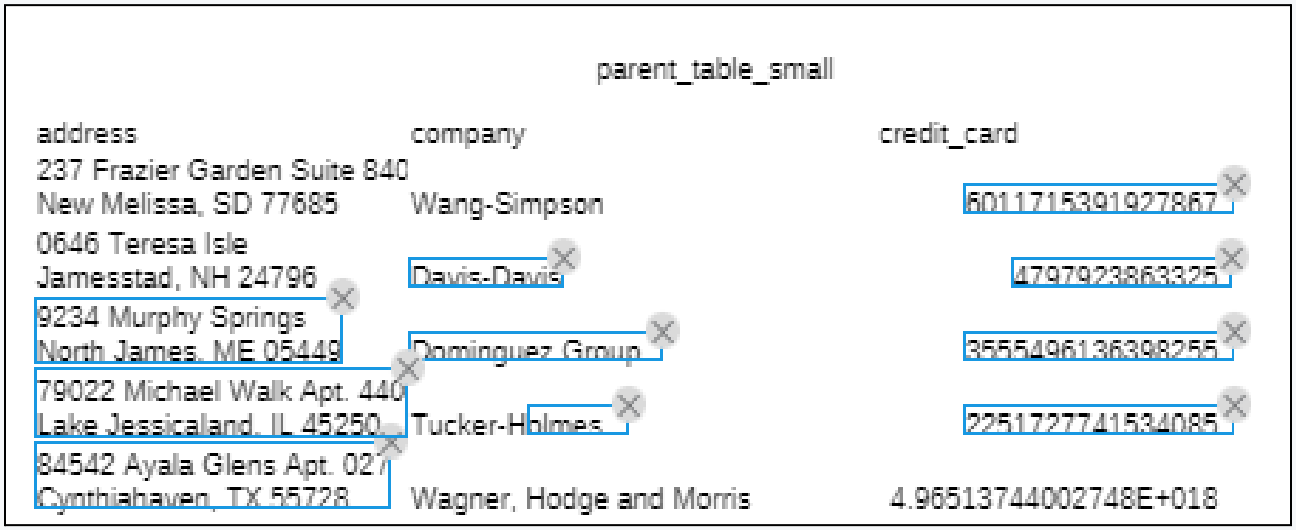
would be redacted. These tickets would appear as if a black marker was used to cross out the information

Redaction

LightBeam can recognize sensitive data in the body and files in Google Drive.

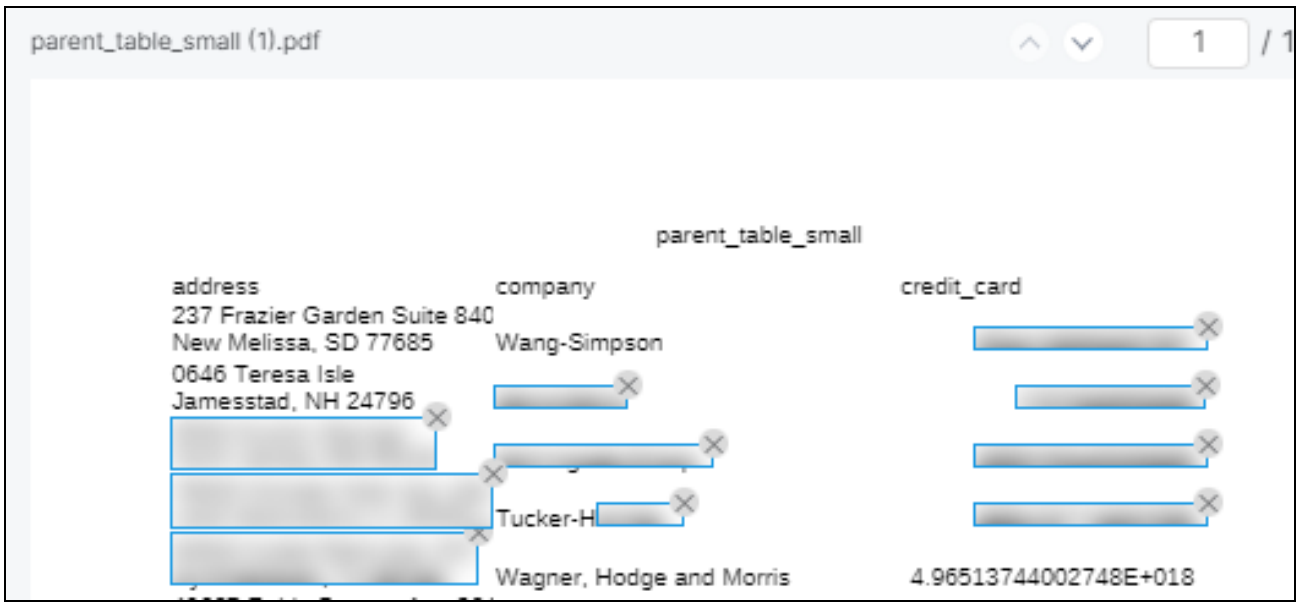
- By scanning and analyzing files and folders, sensitive data selected for redaction is identified.
- The use of masking selectors allows for document-specific redaction rule sets to be turned on or off providing access to the selected attributes.
- PI is obfuscated after the information has met the specified date. The PI would be removed after the retention period has been met and the data is no longer needed.
- Files can safely be archived with all sensitive data redacted.

In the following example, we see a Google Drive folder containing files with sensitive credit card information. LightBeam's Google Drive bot has detected sensitive data in the credit card field of a database table stored in a Google Drive folder.



LightBeam's Google Drive bot has detected sensitive information in the file and has raised an alert to the indicated owner.

After the LightBeam Bot has recognized the sensitive information it has redacted it from the file as configured by the policy.



By establishing Policies, Rule sets, and Permission lists in LightBeam, automated monitoring not only provides visibility of the data inventory, it also continually enforces the data management rules to manage privacy risk in key applications. By having a process to automatically remove PI from all old files, privacy officers have reduced their PI footprint and reduced their risk of inappropriate loss or disclosure.

Connecting Org Google Drive to LightBeam

Connecting any Google Drive data source to LightBeam is a simple 3 step process.

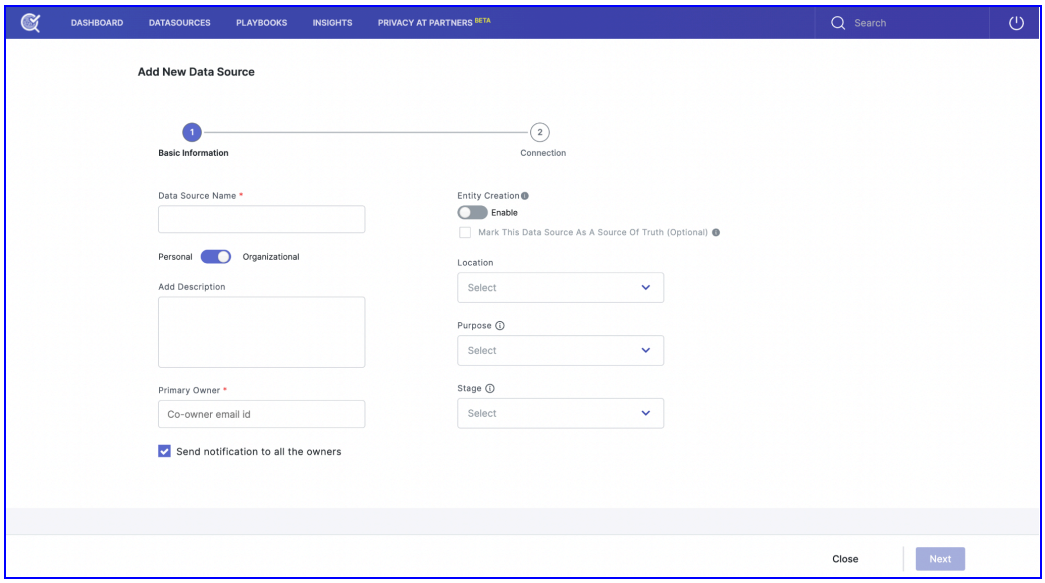
1. Create a new Gdrive data source.
2. Connect to the new Gdrive
3. Configure scan preferences

To begin from the LightBeam dashboard, select DATA SOURCES. From the DataSource home screen select Add Data Source from the top right of the page. From the application list select New Google Drive and continue with the steps below.

Step 1:

Complete the Basic Information page to create a new data source. Keep the following guidance in mind when creating a new data source:

1. Make sure the name is not repeated, as it acts as metadata for the data source and must be unique. E.g., Google_Drive_Sandbox
2. Use the description to explain the kind of information the data source contains. E.g., all HR-related documents stored here
3. LightBeam uses the email ID stated as the Primary Owner to send alert notifications.
4. You can add another email for notifications using the co-owner button. Remember to check the 'send notification to all owners' box.
5. Entity Creation: Enable to create entities out of this data source.
6. Location tells where the data source is located.
7. Purpose tells for what purpose is this data source storing or processing data.
8. Stage tells what stage the data belongs in could be the source, processing, transactional, archival, etc.



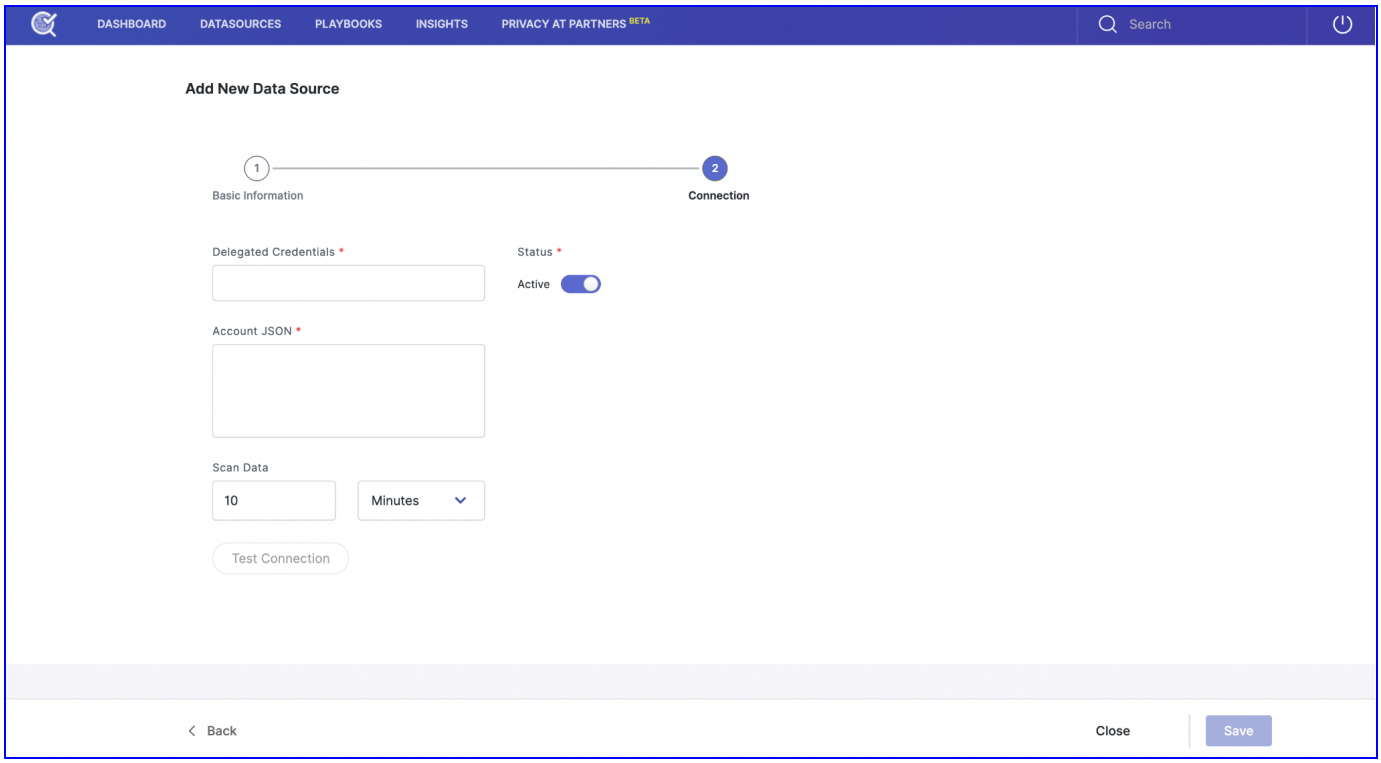
Step 2: Connection Details

1. Select the start date from when you want to start scanning the tickets
2. Add subdomain name
3. Configure API Token for authentication

Add the Email ID from which the API token has been created

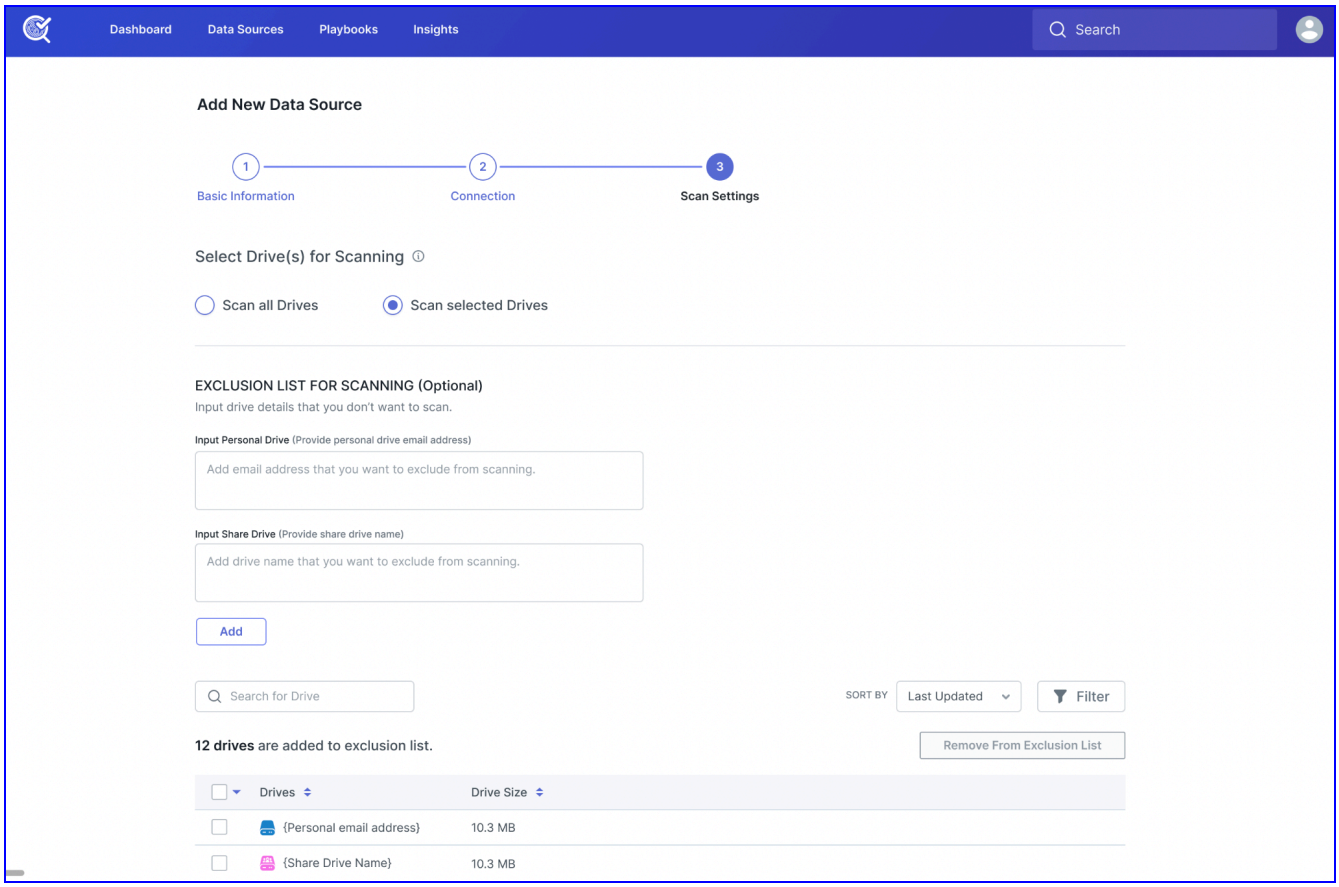
Add API token

You can test the connection and see if the connection went through



Step 3: Scan Settings

1. Scan all drives
2. Scan selected drives
3. Exclusion list
    a. Add personal drives
    b. Add shared drives



Conclusion

Managing the appropriate use of personal information is challenging for any organization. Administrative controls like policies, procedures, and employee training are only as good as their execution which is often an afterthought in many organizations.

By using the power of AI to diligently scan and monitor data applications and repositories, data officers can apply significant technical control over how data is

stored and processed. This in turn provides privacy teams with new accurate pictures of their sensitive data and how it is used in the organization. By understanding all sensitive data in an organization LightBeams 360 degree view allows for more advanced reviews and proactive actions to be taken to manage privacy operations and reduce overall privacy risk.

Appendix

Revision History

Reference Architecture Update	Date	Author
First generally available version.	11/30/2023	Bill Schaumann

About LightBeam

With its focus on Data Privacy Automation (DPA), LightBeam is pioneering a unique identity-centric and automation-first approach to the data privacy and data security markets. Unlike siloed solutions, LightBeam's Data Privacy Automation (DPA) ties together sensitive data discovery, cataloging, access, and data loss

prevention (DLP), and makes the right (sensitive) identity-centric data available to the right people and teams. It becomes the privacy control tower providing a 360-degree view of PII/PHI sensitive data sprawl. LightBeam enables privacy officers to set policies to automate their enforcement, while information security executives can finally rest assured that sensitive data is being used and accessed securely.