CloudLock Whitepaper

Identifying, Classifying, and Protecting Personally Identifiable Information in Google Drive (Docs)

The Enterprise Guide To Securing Sensitive Data In Google Drive

At a Glance

Intended Audience:

- C-level security, collaboration, and trust professionals at companies using or considering the Google Apps productivity suite
- Domain Administrators at organizations using Google Apps

Takeaway

- Readers will learn how to find, classify, and secure PII in Google Drive

In This Whitepaper:

Introduction

As more companies begin using Google Drive (Docs) as their primary cloud file server, they need the ability to identify and protect Personally Identifiable Information (PII) stored and shared in Google Docs.

"Enterprises handling sensitive data, including personally identifiable information, have a responsibility to protect access to this information regardless of whether the information is kept on-premises or in a cloud-based provider. Moving to the cloud doesn't remove this responsibility, nor does it shift it solely to the cloud provider– enterprise security policies must be extended to cloud-based information and we need to be able to show compliance with these policies."

— Neil MacDonald, Vice President and Gartner Fellow

The following guide will give in-depth steps to identify, classify, and secure PII in Google Drive (Docs) including:

- Credit card numbers
- Social Security numbers
- Phone numbers, addresses, and other sensitive data using custom regular expressions for pattern matching

Best Practices for Identifying, Classifying, and Protecting PII in Google Drive

As more companies begin using Google Drive (Docs) as their primary cloud file server, they need the ability to identify and protect Personally Identifiable Information (PII) stored and shared in Google Docs.

"As a publicly traded pharmaceutical company, we face the challenge of complying with SOX and FDA regulations as well as protecting our IP. With CloudLock's pattern matching engine, we're able to identify PII as well as very sensitive data related to our intellectual property, giving us the ability to find, classify, and protect our data in Google Drive."

— *Nathan McBride,*
Vice President of IT, AMAG Pharmaceuticals

Personally Identifiable Information (PII) is information that can be used to uniquely identify, contact, or locate a single individual. A few examples:

- Credit card numbers
- Social Security numbers
- Phone numbers, addresses, and other sensitive data

Many regulations and privacy laws also dictate keeping PII secure:

PCI DSS – The Payment Card Industry Standards Security Council requires organizations that handle payment cards to comply with security standards and protect the private information of bankcard holders during any transaction. This regulation requires that organizations secure all information related to cardholders regardless of the location of the data.

FISMA – The Federal Information Security Management Act requires government agencies to develop information security practices to protect sensitive and private information that support the operations and assets of each agency.

FERPA – The Family Educational Rights and Privacy Act is a federal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education.

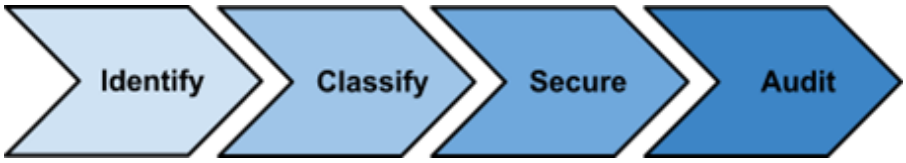
Many other regulations such as SOX and CIPA also require organizations to identify, classify and secure sensitive information to be able to comply with these and many other regulations as well as internal governance and Acceptable Use Policies (AUPs.)

All of the above mentioned regulations clearly indicate that they apply to all data that organizations create, store, and process regardless of media and the location of the data. As more and more information is stored in the cloud, organizations are now required to extend their security practices to data stored in the cloud.

Enterprises must be able to answer the following questions about PII and other sensitive data:

1. How can we ensure PII and PCI compliance in Google Drive (Docs)?
2. How can we find and protect files containing Social Security Numbers and Credit Card Numbers?
3. How do I perform RegEx searches for pattern matching in Google Docs (Drive)?

Addressing compliance regulations for data stored in the cloud should follow the same established framework as all other data. This framework consists of the following steps:



Even though the basic concepts have not changed since the inception of these regulations, the tools used to identify, classify, secure and audit data on premise no longer can be used with data stored in the cloud in general, and Google Drive (Docs) specifically.

When addressing the compliance challenges in cloud data repositories, the following best practices apply:

- Ideally, data should not leave its source system for compliance processing to avoid creating additional copies of sensitive information
- Costs of protecting sensitive data should not negate the cost-savings benefits of moving to the cloud
- Compliance is not a one-time effort. Application(s) chosen for this task need to provide ongoing scanning, classification, alerting, and corrective actions when PII is found
- Compliance and auditing tasks should not distract the organization from performing its core activities, and should not hinder operational efficiency
- End-Users and data owners should be involved in security and compliance tasks. Delegation to departmental security officers and end-users who know the data intimately is necessary to achieve compliance while not burdening IT with all compliance-related tasks

With CloudLock's compliance scan, companies can find and protect PII. The following guides give step-by-step details on how to identify, classify, and protect PII data like credit card number, social security numbers, and other sensitive data in Google Docs and Drive:

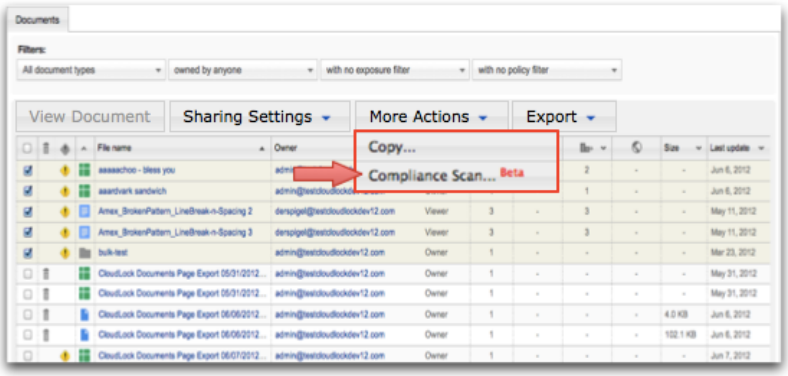
- How To Find and Protect Personally Identifiable Information (PII) In Google Docs and Drive
- How To Find and Protect Credit Card Numbers In Google Docs and Drive
- How To Find and Protect Social Security Numbers In Google Docs and Drive
- How To Find and Protect Sensitive Data with Custom Pattern Matching in Google Docs and Drive

How To Find and Protect Personally Identifiable Information (PII) In Google Docs and Drive

Step #1: Identify the Data Subject to Compliance

Since not all data needs to be examined, identify which data should be scanned for compliance. Some examples of data that should be inspected:

- Files created by certain users
- Data stored in specific collections/ folders
- Files exposed publicly, externally, or internally
- Files by keyword



In the CloudLock Document Browser, select the files to review and select the Compliance Scan (under the More Actions menu).

Step #2: Classify the Data

This step will select the relevant compliance scan to classify the data. CloudLock follows the strict standards put forth by the Social Security Administration and the Payment Card Industry Standards Security Council to perform our pattern recognition.

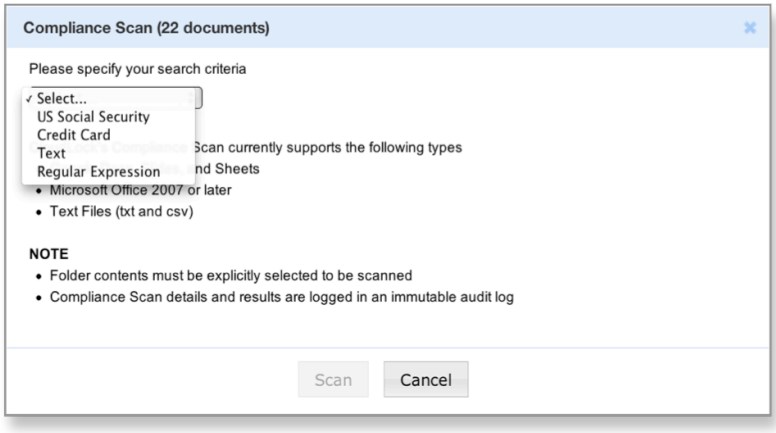
First, choose one of the following options:

- US Social Security – Identify documents containing Social Security numbers
- Credit Card – Identify and flag documents containing credit card numbers
- Custom – Define any custom regular expression patterns (Text or RegEx) relevant to sensitive data.

    For example:

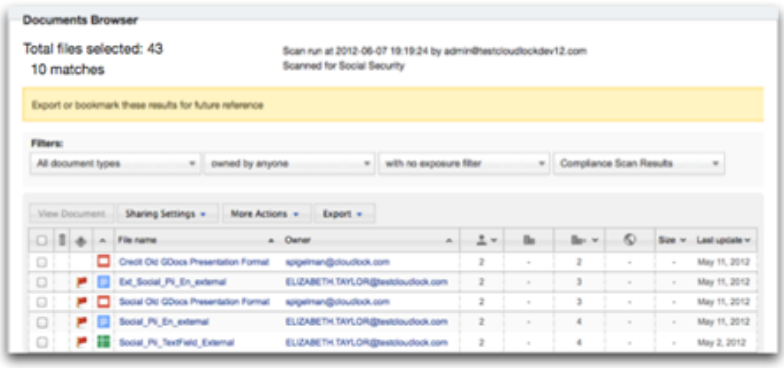
    1. Healthcare Numbers
    2. National ID numbers
    3. Student IDs
    4. Vehicle Registration Numbers
    5. Date of Birth
    6. Genetic Information
    7. Passport
    8. Digital Identification
    9. Postal Codes
    10. Part Numbers

    See the full list of RegEx supported actions.



When the scan completes, a report classifies documents by status:

- Pattern found
- Scan error – document was not scanned
- Unsupported document format, not scanned



Step #3: Investigate and Secure Flagged Documents

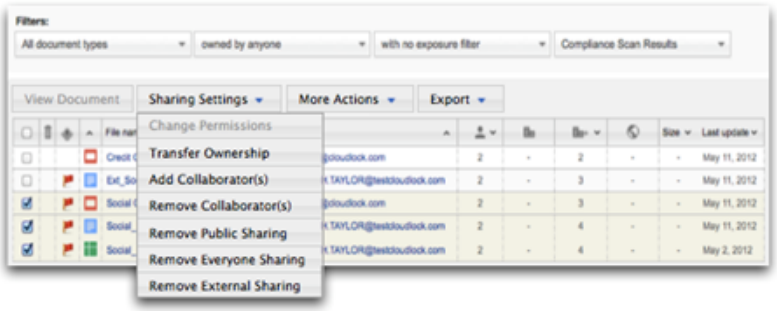
Now that documents have been found matching the pattern selected, open the documents to verify that they contain sensitive data. To do this, click on the document title, and then click “View document” or “View spreadsheet/form”. Viewing the document will grant you view permissions, and the action will be recorded in the audit log.

First Name	Last Name	Birth Date	SSN
Todd	Davis	1/17/1968	457-55-5469
Hilda	Whitcher	3/16/1907	219-09-9999
Douglas	Patterson	8/1/1900	078-05-1120
Sam	Elliot	5/12/1984	123-80-0909
Thurston	Howell	6/9/1979	312-45-7634
Mimi	Beardsley	12/5/1965	102-78-9431
Travis	Milton	6/16/1956	898-21-2964
Saleh	Wintumbah	8/16/1956	103-73-3986

Depending on the sensitivity of the document and how it is shared, it may be necessary to remove collaborators, change access rights, remove public or external exposure, and notify data owners.

CloudLock enables customers to secure documents individually or in bulk for any of the following actions:

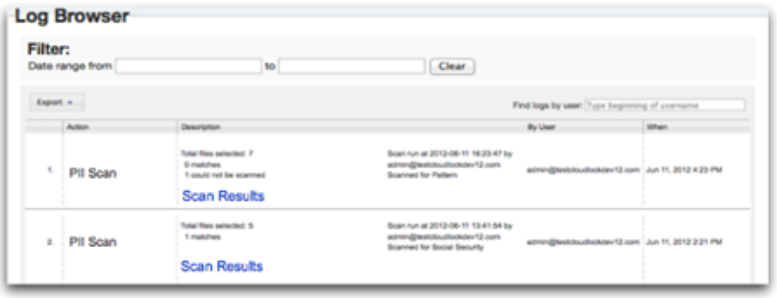
- Add or remove collaborators
- Remove public, external, or internal exposure
- Copy files and transfer ownership
- Change permissions



Step #4: Audit

All actions performed in CloudLock are recorded in an immutable audit log and can be used as evidence of compliance to regulations and internal policies. After addressing documents flagged for compliance issues, completing a new scan will produce a new audit entry, confirming that the issue has been resolved.

The CloudLock audit log is instrumental in demonstrating that compliance is not just a one-time effort, but rather an ongoing practice that every organization should adopt as part of their regular operational procedures.

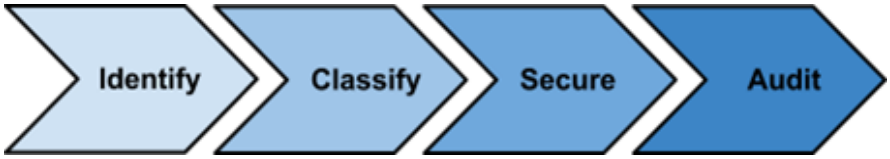


Keeping PCI information secure is also dictated by regulations and privacy laws including:

> PCI DSS – The Payment Card Industry Standards Security Council requires organizations that handle payment cards to comply with security standards and protect the private information of bank card holders during any transaction. This regulation requires that organizations secure all information related to cardholders regardless of the location of the data.

PCI DSS clearly applies to credit card data that organizations create, store, and process regardless of media and the location of the data. As more information is stored in the cloud, organizations are now required to extend their security practices to data stored in the cloud.

Protecting credit card numbers in data stored in the cloud should follow the same established framework as all other data. This framework consists of the following steps:



Challenges In Securing PII In The Cloud

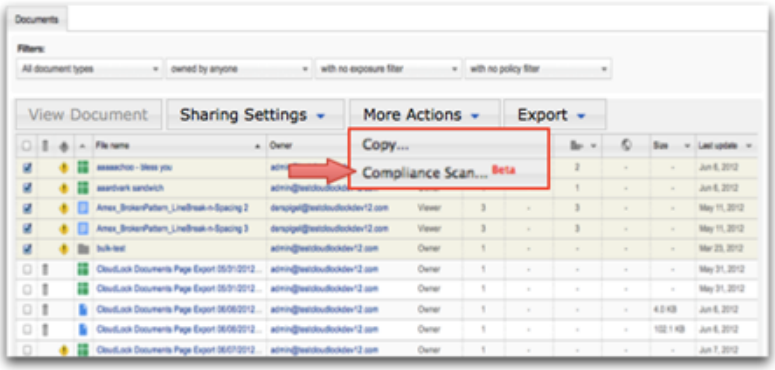
The tools used to identify, classify, secure and audit credit card data on-premise no longer can be used with data stored in the cloud in general, and Google Drive (Docs) specifically.

With CloudLock Compliance Scan, companies can find and protect PII. The following guide will show how to use CloudLock to find PCI info in Google Drive (Docs).

Step #1: Identify the Data Subject to Compliance

Since not all data needs to be examined, identify which data should be scanned for PCI data. Some examples of data that should be inspected:

- Files created by certain users
- Data stored in specific collections/ folders
- Files exposed publicly, externally, or internally
- Files by keyword

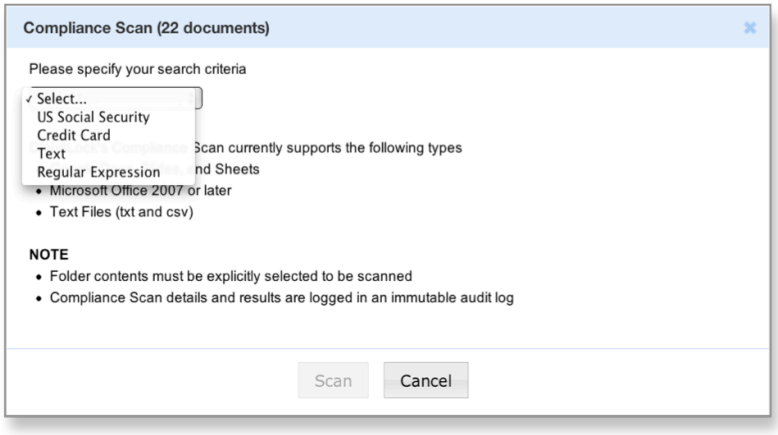


In the CloudLock Document Browser, select the files to review and select the Compliance Scan (under the More Actions menu)

Step #2: Classify the Data

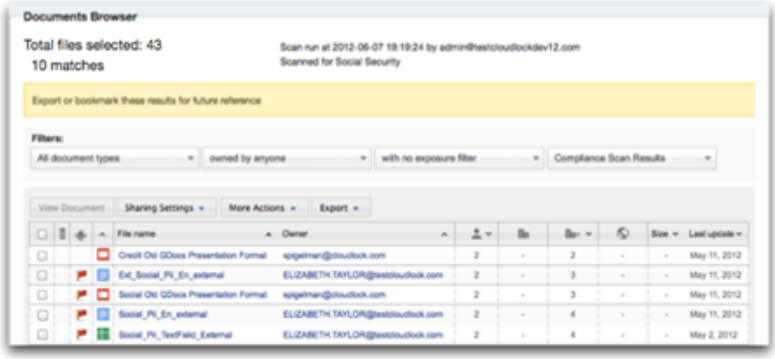
This step will select the relevant compliance scan to classify the data. CloudLock follows the strict standards put forth by the Payment Card Industry Standards Security Council to perform our pattern recognition.

First, choose the “Credit Card” option



When the scan completes, a report classifies documents by status:

- Pattern found
- Scan error – document was not scanned
- Unsupported document format, not scanned



Step #3: Investigate and Secure Flagged Documents

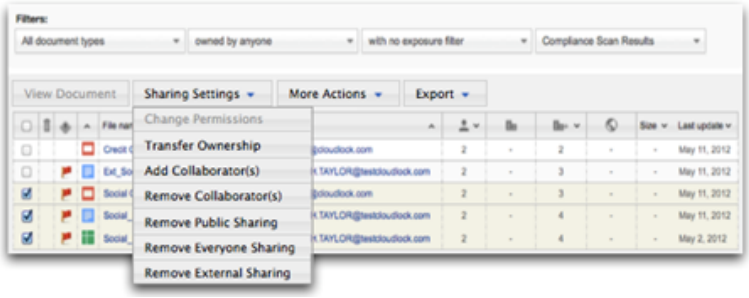
Now that documents have been found matching the pattern selected, open the documents to verify that they contain sensitive data. To do this, click on the document title, and then click “View document” or “View spreadsheet/form”. Viewing the document will grant you view permissions, and the action will be recorded in the audit log.

First Name	Last Name	Credit Card Type	Credit Card Number
Clyde	Fretwell	Visa	4125 3095 8907 2856
Elisha	Skinner	MasterCard	3900 6811 0987 2321
Rosa	Scott	American Express	3400 0000 0000 009
Thomas	Cooke	Discover	6011 0000 0000 0004
Julia	Blanchard	Visa	4300 0915 0065 3831
Page	Harrell	Visa	9316 0132 1432 1298
Herbert	Modlin	Visa	9300 1254 0909 3235
Mattie	Phelps	MasterCard	1900 6811 0987 2321
Celia	Harmon	American Express	3400 0000 1112 009
Thomas	Godwin	Visa	9212 0987 2412 0987

Depending on the sensitivity of the document and how it is shared, it may be necessary to remove collaborators, change access rights, remove public or external exposure, and notify data owners.

CloudLock enables customers to secure documents individually or in bulk for any of the following actions:

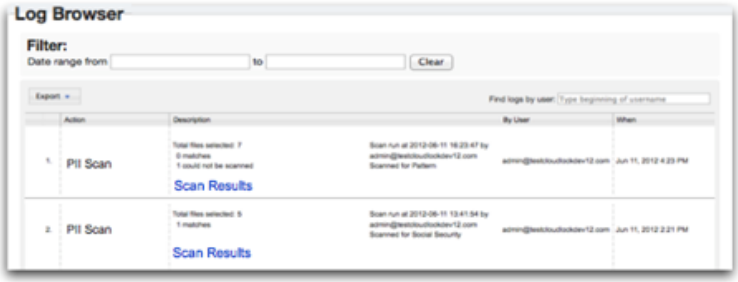
- Add or remove collaborators
- Remove public, external, or internal exposure
- Copy files and transfer ownership
- Change permissions



Step #4: Audit

All actions performed in CloudLock are recorded in an immutable audit log and can be used as evidence of compliance to regulations and internal policies. After addressing documents flagged for compliance issues, completing a new scan will produce a new audit entry, confirming that the issue has been resolved.

The CloudLock audit log is instrumental in demonstrating that compliance is not just a one-time effort, but rather an ongoing practice that every organization should adopt as part of their regular operational procedures.



Keeping SSN information secure is also dictated by regulations and privacy laws including:

> FISMA – The Federal Information Security Management Act requires government agencies to develop information security practices to protect sensitive and private information that support the operations and assets of each agency.
>
> Other regulations such as SOX and CIPA also require organizations to identify, classify and secure sensitive information to be able to comply with these and other regulations as well as internal governance and Acceptable Use Policies (AUPs):

These regulations apply to social security number data that organizations create, store, and process regardless of media and the location of the data. As more information is stored in the cloud, organizations are now required to extend their security practices to data stored in the cloud.

Protecting social security numbers in data stored in the cloud should follow the same established framework as all other data. This framework consists of the following steps:

Challenges In Securing PII In The Cloud

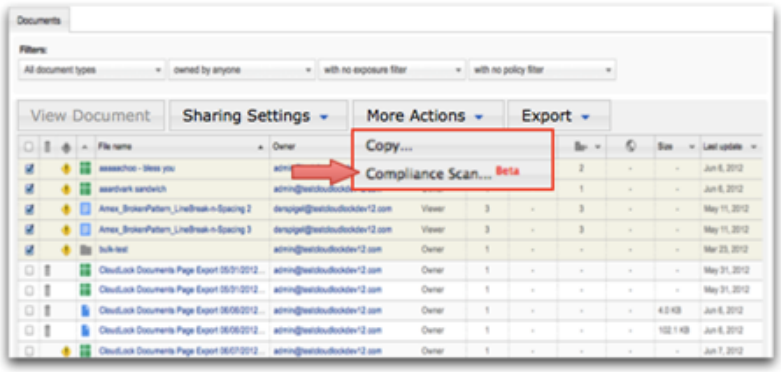
Though the basic concepts have not changed since the inception of data privacy regulations, the tools used to identify, classify, and secure social security number data on premise no longer can be used with data stored in the cloud in general, and Google Drive (Docs) specifically.

With CloudLock Compliance Scan, companies can find and protect social security number information. The following guide will show how to use CloudLock to find social security numbers in Google Drive (Docs).

Step #1: Identify the Data Subject to Compliance

Since not all data needs to be examined, identify which data should be scanned for SSN data. Some examples of data that should be inspected:

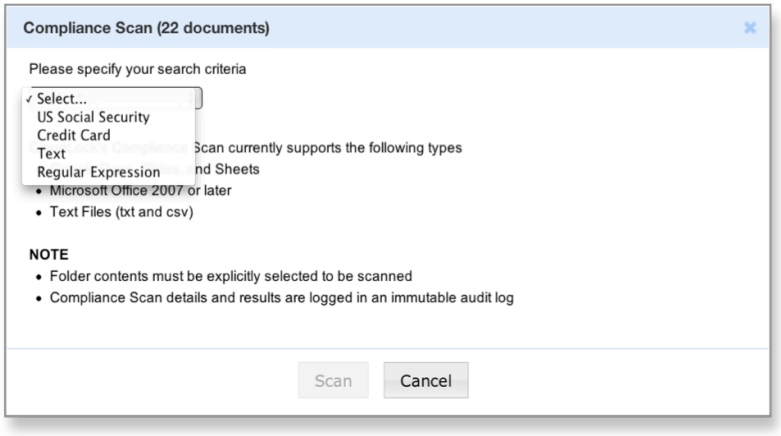
- Files created by certain users
- Data stored in specific collections/ folders
- Files exposed publicly, externally, or internally
- Files by keyword



In the CloudLock Document Browser, select the files to review and select the Compliance Scan (under the More Actions menu)

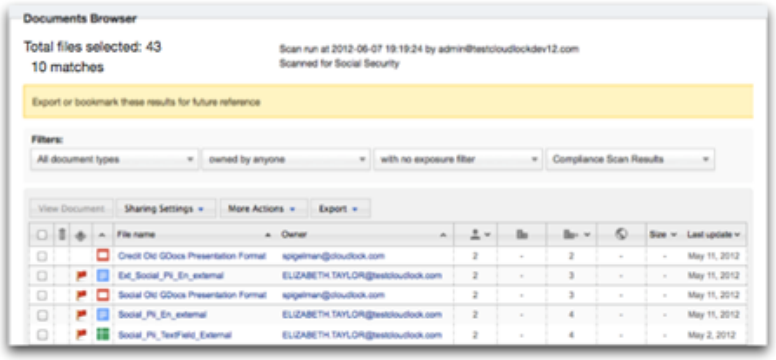
Step #2: Classify the Data

This step will select the relevant compliance scan to classify the data. CloudLock follows the strict standards put forth by the Social Security Administration to perform our pattern recognition. First, choose the “US Social Security” option



When the scan completes, a report classifies documents by status:

- Pattern found
- Scan error – document was not scanned
- Unsupported document format, not scanned



Step #3: Investigate and Secure Flagged Documents

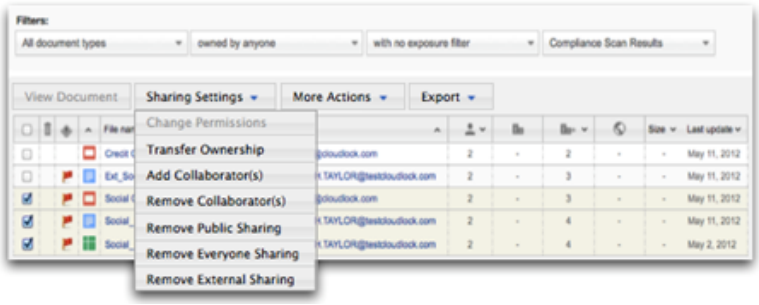
Now that files have been found with social security information, open the documents to inspect the contents. To do this, click on the document title, and then click “View document” or “View spreadsheet/form”. Viewing the document will grant you view permissions, and the action will be recorded in the audit log.

First Name	Last Name	Birth Date	SSN
Todd	Davis	1/17/1968	457-55-5469
Hilda	Whitcher	3/16/1907	219-09-9999
Douglas	Patterson	8/1/1900	078-05-1120
Sam	Elliot	5/12/1984	123-80-0909
Thurston	Howell	6/9/1979	312-45-7634
Mimi	Beardsley	12/5/1965	102-78-9431
Travis	Milton	6/16/1956	898-21-2964
Saleh	Wintumbah	8/16/1956	103-73-3986

Depending on the sensitivity of the document and how it is shared, it may be necessary to remove collaborators, change access rights, remove public or external exposure, and notify data owners.

CloudLock enables customers to secure documents individually or in bulk for any of the following actions:

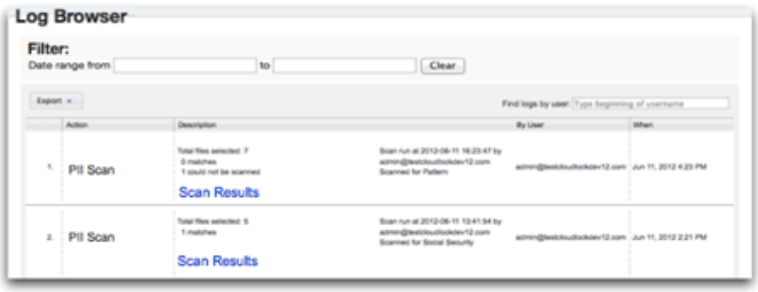
- Add or remove collaborators
- Remove public, external, or internal exposure
- Copy files and transfer ownership
- Change permissions



Step #4: Audit

All actions performed in CloudLock are recorded in an immutable audit log and can be used as evidence of compliance to regulations and internal policies. After addressing documents flagged for SSN data, completing a new scan will produce a new audit entry, confirming that the issue has been resolved.

The CloudLock audit log is instrumental in demonstrating that compliance is not just a one-time effort, but rather an ongoing practice that every organization should adopt as part of their regular operational procedures.



How To Find and Protect Sensitive Data with Custom Pattern Matching in Google Docs and Drive

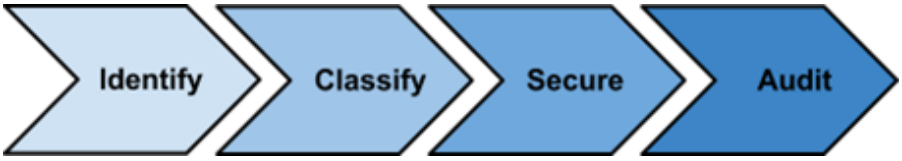
As companies begin using Google Drive (Docs) as their primary cloud file server, they need the ability to identify and protect sensitive data stored and shared in Google Docs.

When most people hear the term 'sensitive information', they think of Personally Identifiable information (PII) like social security numbers, and Payment Credit Industry (PCI) data like credit card numbers. But in addition to these examples, the term "sensitive information" means different things to different organizations.

Some examples:

Type of Organization	Sensitive Information	Compliance Need
e-Commerce	Customer Credit Card Information	PII
Product Development	Intellectual Property, Patents, Trade Secrets	Internal Governance
Publicly Traded	Financial Information	SOX
Hosting Customer Data	Sensitive Customer Data	Sensitive customer data Compliance requirements and legal liabilities of customers
Government	Government As defined by NIST Guide to Protecting the Confidentiality of Personally Identifiable Information (SP 800-122)	FISMA
Higher Education	Higher Ed Employee PII, Student Social Security Numbers	Data Privacy Regulations
K-12	K-12 Students information / PII Employee PII	FERPA, CIPA

Protecting sensitive corporate data stored in the cloud should follow the same established framework as all other data. This framework consists of the following steps:

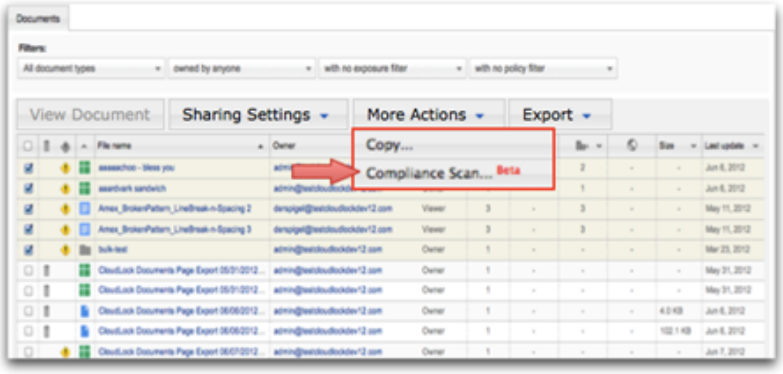


With CloudLock's compliance scan, companies can find and protect sensitive information using regular expressions and pattern matching. The following guide will show how to use CloudLock to find sensitive data with custom pattern matching in Google Docs and Drive.

Step #1: Identify the Data To Be Searched

Since not all data needs to be examined, identify which data should be scanned for pattern matching. Some examples of data that should be inspected:

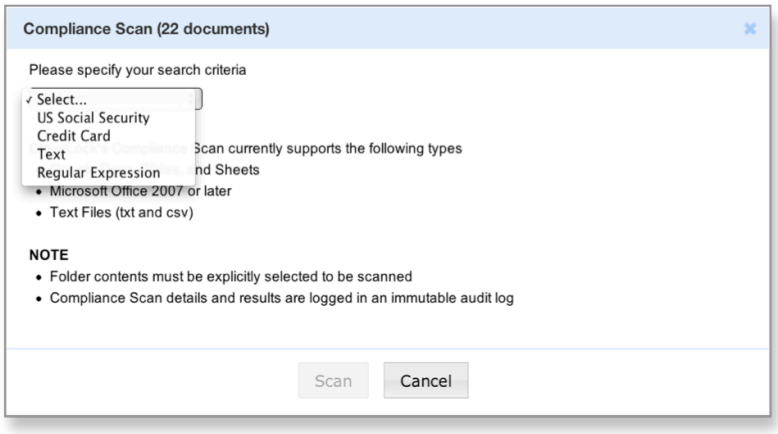
- Files created by certain users
- Data stored in specific collections/ folders
- Files exposed publicly, externally, or internally
- Files by keyword



In the CloudLock Document Browser, select the files to review and select the Compliance Scan (under the More Actions menu)

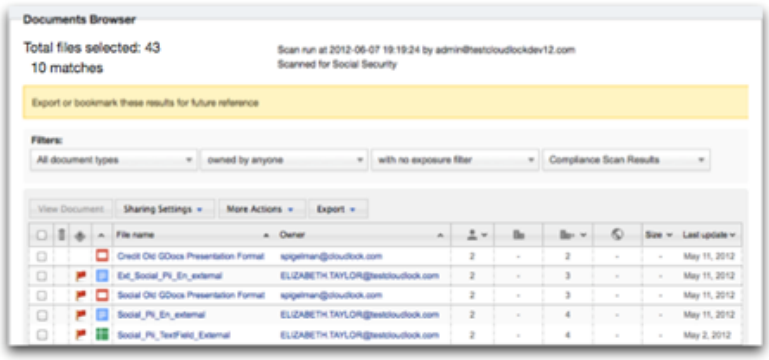
Step #2: Classify the Data

This step will select the relevant compliance scan to classify the data. First, choose either the “Text” or “Regular Expression” option



When the scan completes, a report classifies documents by status:

- Pattern found
- Scan error – document was not scanned
- Unsupported document format, not scanned



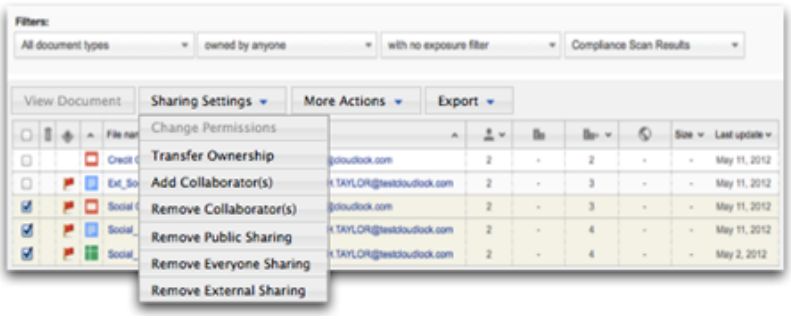
Step #3: Investigate and Secure Flagged Documents

Now that files have been found matching the pattern selected, open the documents to inspect the contents. To do this, click on the document title, and then click “View document” or “View spreadsheet/form”. Viewing the document will grant you view permissions, and the action will be recorded in the audit log.

Depending on the sensitivity of the document and how it is shared, it may be necessary to remove collaborators, change access rights, remove public or external exposure, and notify data owners.

CloudLock enables customers to secure documents individually or in bulk for any of the following actions:

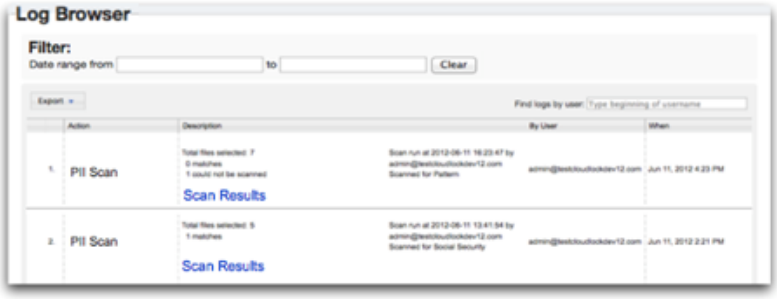
- Add or remove collaborators
- Remove public, external, or internal exposure
- Copy files and transfer ownership
- Change permissions



Step #4: Audit

All actions performed in CloudLock are recorded in an immutable audit log and can be used as evidence of compliance to regulations and internal policies. After addressing documents flagged by the pattern-matching engine, completing a new scan will produce a new audit entry, confirming that the issue has been resolved.

The CloudLock audit log is instrumental in demonstrating that compliance is not just a one-time effort, but rather an ongoing practice that every organization should adopt as part of their regular operational procedures.



Summary:

Regardless of the type and location of Personally Identifiable Information, companies are required to identify, classify, secure, and audit PII data. With CloudLock Compliance Scan, organizations using Google Docs have the ability to keep very sensitive information safe and use Google Drive as a cloud file server.

About CloudLock

CloudLock helps enterprises extend their data security practices and policies to the cloud. CloudLock's suite of security applications give businesses the controls and visibility they need to take advantage of the collaboration benefits of public cloud offerings, without sacrificing on security. The largest Google Apps customers in the world trust CloudLock to secure their data. For more information about the company or reseller opportunities call (781) 996-4332 or visit http://www.cloudlock.com/.