Coveo Platform 7.0

Google Sites Connector Guide

Notice

The content in this document represents the current view of Coveo as of the date of publication. Because Coveo continually responds to changing market conditions, information in this document is subject to change without notice. For the latest documentation, visit our website at www.coveo.com.

Document part number:	PM-131219-EN
Publication date:	1/3/2019

Table of Contents

1. Google Sites Connector

CES 7.0.6225+ (December 2013)

The Coveo beta connector for Google Sites allows Coveo administrators to index and integrate the content of one or more domain or private Google Sites into the Coveo unified index so that in the Coveo search interfaces, a user can easily find content to which he has access in Google Sites.

1.1 Features

The Google Sites connector features are:

Content indexing

The connector can index Google Sites content from either a single user Google Account or a Google Apps domain.
The connector can index the following Google Sites content type to make it searchable by end-users:

- Web pages
- Attachments
- Announcements (Announcement pages)
- Files and web attachments (File cabinets)
- List items (List pages)

Note: The comments cannot be indexed because the current Google Sites API only allows to access comments created with the old commenting systems, not those created with the new commenting system.

Site-level permission indexing

The connector can index the Google Sites site-level permissions for each item. This means that, in Coveo search interfaces, a user searching Google Sites content only sees the content to which he has access as specified in site-level permissions.

Note: The connector cannot index page-level permissions because this information is not available through the Google Sites API. This means that a user that is denied access to a specific Google Sites page could see this page in Coveo search results.

Incremental refresh

Supports incremental refresh to periodically query Google Sites for the latest edits, keeping the index content up-to-date.

Note: The incremental refresh has the following limitations:

- Permission changes alone cannot be detected.
- Does not work when the last refresh date is more than 30 days, because list of deleted items are only kept for 30 days.

Multithreading

The connector can run multiple threads, which can improve performances considerably.

Feature History

CES version	Monthly release	Features
7.0.7183+	November 2014	Connector provides default field set and mapping file [more]
7.0.6225+	December 2013	Connector introduction (Beta)

What's Next?

Review the steps to deploy the Google Sites connector (see "Google Sites Connector Deployment Overview" on page 3).

2. Google Sites Connector Deployment Overview

The following procedure outlines the steps needed to deploy the Google Sites connector. The steps indicate the order in which you must perform configuration tasks on both the Google and Coveo servers.

To deploy the Google Sites connector

1. Validate that your environment meets the requirements (see "Google Sites Connector Requirements" on page 6).

2. On the Google console, authorize the Coveo connector to access your Google Sites.

   You must create a Google API Console project to grant the connector access to the Google Sites that you want to index (see "Granting the Connector Access to Your Google Sites" on page 7).

3. On the Coveo server, in the Coveo Administration Tool:

   a. Configure a user identity

      The Coveo connector needs an account to crawl your Google Sites content. For this purpose, select an existing Google Apps account or create a new one. When you want to index permissions, the crawling account must have administrator rights because ACL Feed access is required. When you do not want to index permissions, an account that has read only permissions to the whole content that you want to index is sufficient.

      Create a CES user identity that contains the credentials of your Google Apps crawling account (see "Adding a User Identity" on page 10).

   b. Optionally create security providers

      When you want to index Google Sites site-level permissions, you must create two security providers to get permissions and, resolve and expand groups.

      In Google Sites, users are identified by their email addresses. Consequently, permissions returned by the Google Apps security provider for each document are email addresses. The Google Apps security provider then requires another security provider to uniquely identify users from their email addresses.

      i. Start by selecting or creating an Email or an Active Directory security provider that the Google Apps security provider will use to resolve and expand groups. The security provider type to use depends on how users are authenticated when they access the search interface:

         - When authenticated with their email address, use an Email security provider with no user identity (see "Configuring an Email Security Provider" on page 14).

         - When authenticated with an Active Directory account, use an Active Directory security provider (see "Configuring an Active Directory Security Provider" on page 15).

Notes:

- An Active Directory security provider is appropriate only when the User Principal Name (UPN) matches the email address for all users.
- CES comes with an Active Directory security provider that you can configure to connect to the default domain. When your environment contains more than one domain, you can select an Active Directory security provider that you created for other domains.

ii. Create a Google Apps security provider that the connector uses to resolve indexed permissions (see "Configuring a Google Sites Security Provider" on page 12).

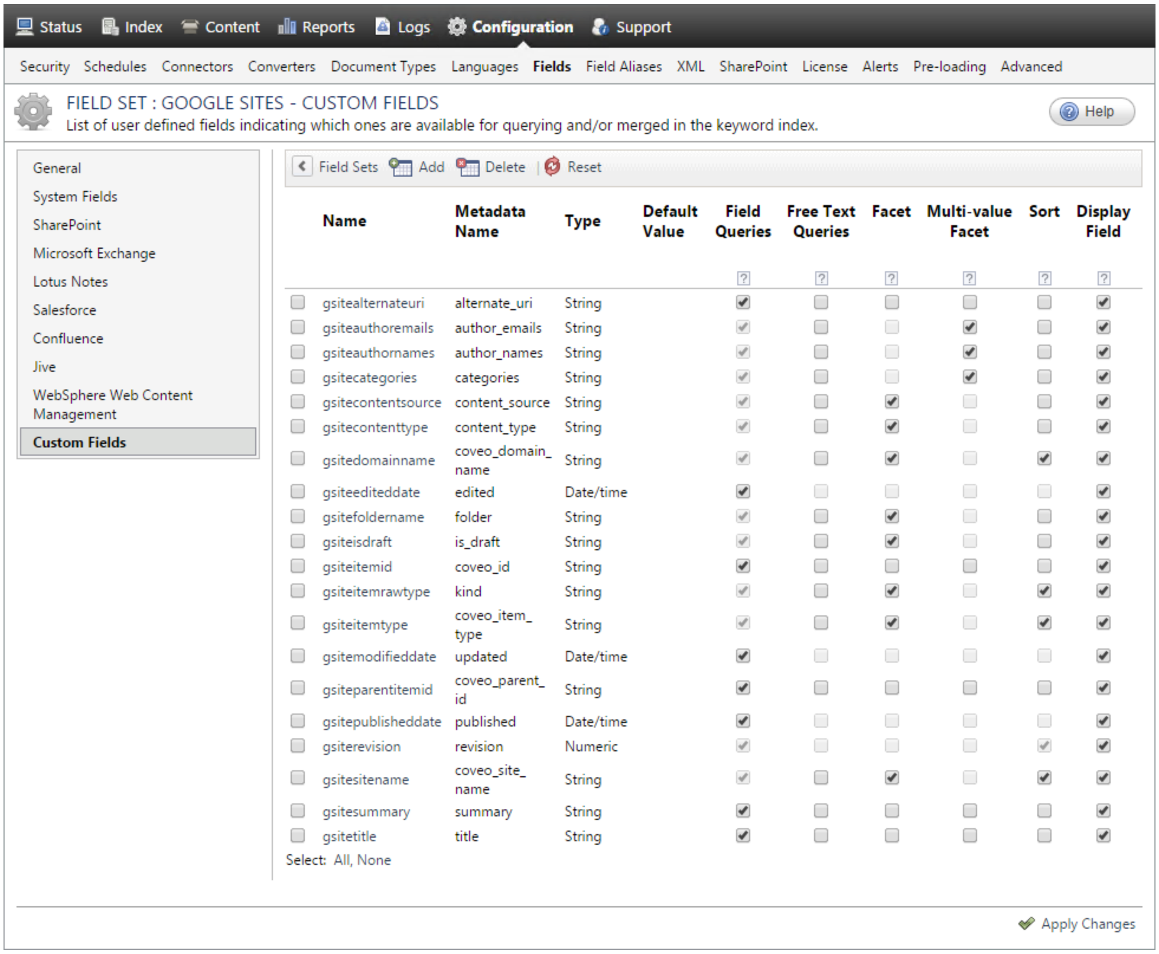
c. CES 7.0.7104– (October 2014) Optionally, when you want to index Google Sites metadata, create a custom mapping file.

A sample mapping file defines a few custom Google Sites field that can be used to create a rich Google Sites search interface, for example, providing more specific facets (see "Creating a Custom Google Sites Connector Mapping File" on page 18).

Note: CES 7.0.7183+ (November 2014) The Google Sites connector comes with a default mapping file (see Configuring and Indexing a Google Sites Source).

d. Create a Google Sites field set to take advantage of the available Google Sites metadata.

It is recommended to start by importing the default Google Sites field set file (`[CES_Path]\Bin\Coveo.CES.CustomCrawlers.GoogleSites.FieldSet.xml` to be able to easily add Google Sites specific facets to your Coveo search interfaces .



FIELD SET : GOOGLE SITES - CUSTOM FIELDS

List of user defined fields indicating which ones are available for querying and/or merged in the keyword index.

	Name	Metadata Name	Type	Default Value	Field Queries	Free Text Queries	Facet	Multi-value Facet	Sort	Display Field
☐	gsitealternateuri	alternate_uri	String		☑	☐	☐	☐	☐	☑
☐	gsiteauthoremails	author_emails	String		☑	☐	☐	☑	☐	☑
☐	gsiteauthornames	author_names	String		☑	☐	☐	☑	☐	☑
☐	gsitecategories	categories	String		☑	☐	☐	☑	☐	☑
☐	gsitecontentsource	content_source	String		☑	☐	☑	☐	☐	☑
☐	gsitecontenttype	content_type	String		☑	☐	☑	☐	☐	☑
☐	gsitedomainname	coveo_domain_name	String		☑	☐	☑	☐	☑	☑
☐	gsiteediteddate	edited	Date/time		☑	☐	☐	☐	☐	☑
☐	gsitefoldername	folder	String		☑	☐	☑	☐	☐	☑
☐	gsiteisdraft	is_draft	String		☑	☐	☑	☐	☐	☑
☐	gsiteitemid	coveo_id	String		☑	☐	☐	☐	☐	☑
☐	gsiteitemrawtype	kind	String		☑	☐	☑	☐	☑	☑
☐	gsiteitemtype	coveo_item_type	String		☑	☐	☑	☐	☑	☑
☐	gsitemodifieddate	updated	Date/time		☑	☐	☐	☐	☐	☑
☐	gsiteparentitemid	coveo_parent_id	String		☑	☐	☐	☐	☐	☑
☐	gsitepublisheddate	published	Date/time		☑	☐	☐	☐	☐	☑
☐	gsiterevision	revision	Numeric		☑	☐	☐	☐	☑	☑
☐	gsitesitename	coveo_site_name	String		☑	☐	☑	☐	☑	☑
☐	gsitesummary	summary	String		☑	☐	☐	☐	☐	☑
☐	gsitetitle	title	String		☑	☐	☐	☐	☐	☑

Select: All, None

Note: CES 7.0.7104– (October 2014) It is recommended to download the zipped Google Sites sample Field set XML file, and import it in CES.

The sample Google Sites field set defines the custom fields that match the custom fields defined in the sample mapping file.

e. Configure and index a Google Sites source.

The connector must know details about the authorized access to the Google Sites of your users to index its content (see "Configuring and Indexing a Google Sites Source" on page 21).

f. If you encounter issues, verify if modifying the default value of hidden source parameters can help resolve the problems (see "Modifying Hidden Google Sites Source Parameters" on page 27).

3. Google Sites Connector Requirements

Your environment must meet the following requirements to be able to use the Google Sites connector:

- CES 7.0.6225+ (December 2013)
- Coveo license for the Google Sites connector

  Your Coveo license must include support for the Google Sites connector to be able to use this connector.
- A valid Google Account

  Using a Google Account with administrator privilege, you must log in the Google API Console to authorize the Coveo connector to access your Google Sites (see Granting the Connector Access to Your Google Sites).

What's Next?

Grant Coveo access to your Google Sites (see Granting the Connector Access to Your Google Sites).

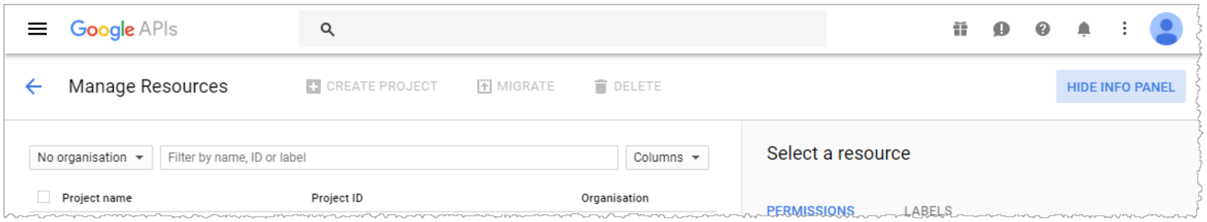
4. Granting the Connector Access to Your Google Sites

Before you can configure Google Sites sources, you must grant the Connector access to the Google Sites content to index. The steps are the same whether you are authorizing access to Google Sites associated to a personal Google account or a Google Apps domain account.

A Coveo source needs to know the values for the Client ID, Client Secret, and OAuth2 Refresh Token associated with your Google Sites.

4.1 Getting Google Sites Client ID and Client Secret values

1. Go to the Google Developers Console, and log in using a Google Account with administrator credentials.



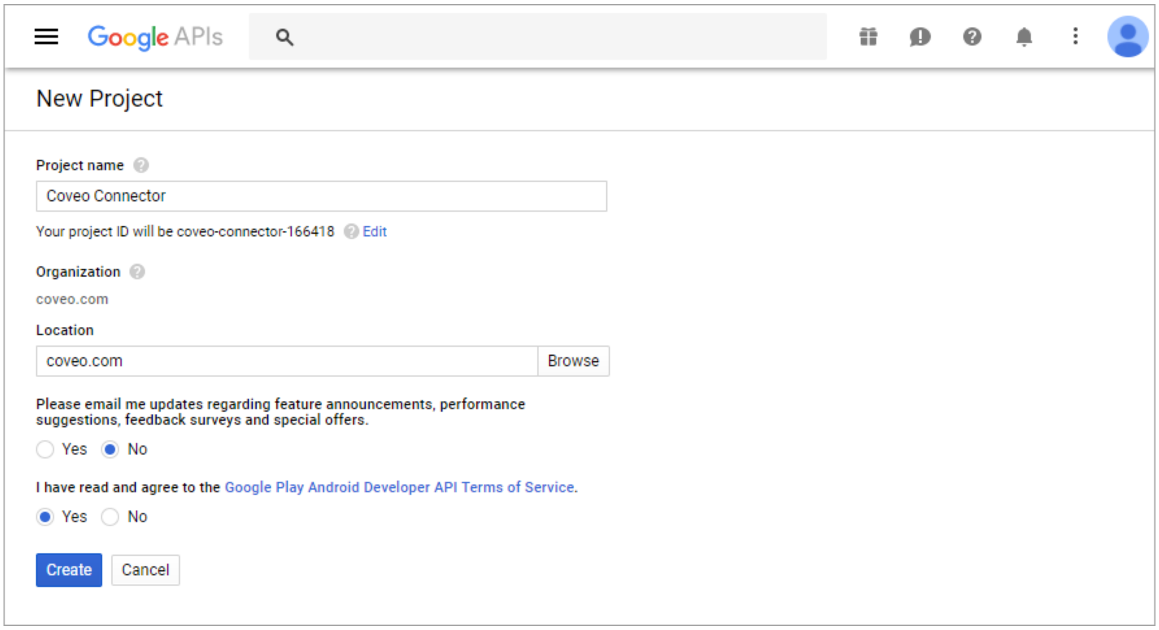
2. At the left of the **Filter by name, ID or label** input, click the drop-down menu, and then select the organization in which you want to create the Google Developer Console project.

3. Create an API project for the Coveo connector (CES 7) or source (Coveo Cloud):

    a. In the **Manage resources** panel, click **Create a project**.

    b. (When your project limit is exceeded) In the **Increase Project Limit** page, click **Request increase**, and then complete the form.

    c. In the **New Project** dialog page, enter the project required information.



New Project

Project name

Coveo Connector

Your project ID will be coveo-connector-166418 Edit

Organization

coveo.com

Location

coveo.com Browse

Please email me updates regarding feature announcements, performance suggestions, feedback surveys and special offers.

○ Yes ● No

I have read and agree to the Google Play Android Developer API Terms of Service.

● Yes ○ No

Create Cancel

i. Enter a **Project name**.

> **Note:** The project ID is automatically created based on the project name. You can always modify the project ID by clicking **Edit**.

ii. (When you create the first project in your organization only) Answer the **Please email me updates regarding feature announcements, performance suggestions, feedback surveys and special offers.** question using the **Yes** or **No** checkbox.

iii. (When you create the first project in your organization only) After you **have read and agree to the Google Play Android Developer API Terms of Service**, click the **Yes** check box.

iv. Click **Create**.

4. Create a Client ID for the Coveo connector.

    a. In the sidebar on the left, select **Credentials**.

    b. In the **Credentials** page, click the **Create credentials** drop-down list menu, and then select **OAuth client ID**.

    c. In the **Create client ID** dialog box:

        a. Under **Application type**, click the **Other** checkbox.

        b. In the box, enter an application **Name**.

        c. Click **Create**.

5. In the **OAuth client** dialog box that appears, note the **Client ID** and **Client secret** values.

4.2 Getting a Google Sites OAuth2 Refresh Token

Now that the Google Sites Connector is registered by creating an API Project and obtained its OAuth2 Client ID, you need to get authorization for the Connector to actually access the Google Sites content.

The Coveo connector needs an OAuth2 refresh token that can only be obtained programmatically (see Using OAuth 2.0 to Access Google APIs).

If you do not have a tool to retrieve the OAuth2 refresh token, contact Coveo Support for assistance.

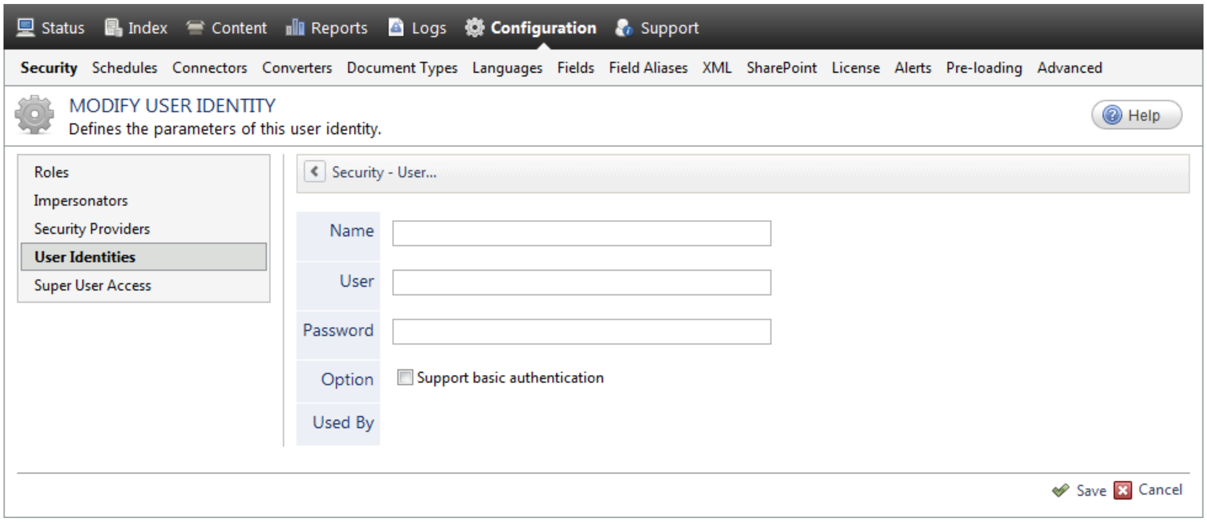
5. Adding a User Identity

A user identity is a set of credentials for a given repository or system that you enter once in CES and can then associate with one or more sources or security providers.

A user identity typically holds the credentials of an account that has read access to all the repository items that you want to index. It is a best practice to create an account to be used exclusively by the Coveo processes and for which the password does not change. If the password of this account changes in the repository, you must also change it in the CES user identity.

To add a user identity

1. On the Coveo server, access the Administration Tool.
2. In the Administration Tool, select **Configuration** > **Security**.
3. In the navigation panel on the left, click **User Identities**.
4. In the **User Identities** page, click **Add**.
5. In the **Modify User Identity** page:



   a. In the **Name** box, enter a name of your choice to describe the account that you selected or created in the repository to allow CES to access the repository.

   Note: This name appears only in the Coveo Administration Tool, in the **Authentication** or **User Identity** drop-down lists, when you respectively define a source or a security provider.

   b. In the **User** box, enter the username for the account that you selected or created to crawl the repository content that you want to index.

   c. In the **Password** box, enter the password for the account.

   d. In the **Options** section, the **Support basic authentication** check box is deprecated and not applicable for most types of repositories. You should select it only when you need to allow CES to send the username and password as unencrypted text.

e. Click **Save**.

Important: When you use Firefox to access the Administration Tool and it proposes to remember the password for the user identity that you just created, select to never remember the password for this site to prevent issues with automatic filling of username and password fields within the Coveo Administration Tool.

6. Configuring a Google Sites Security Provider

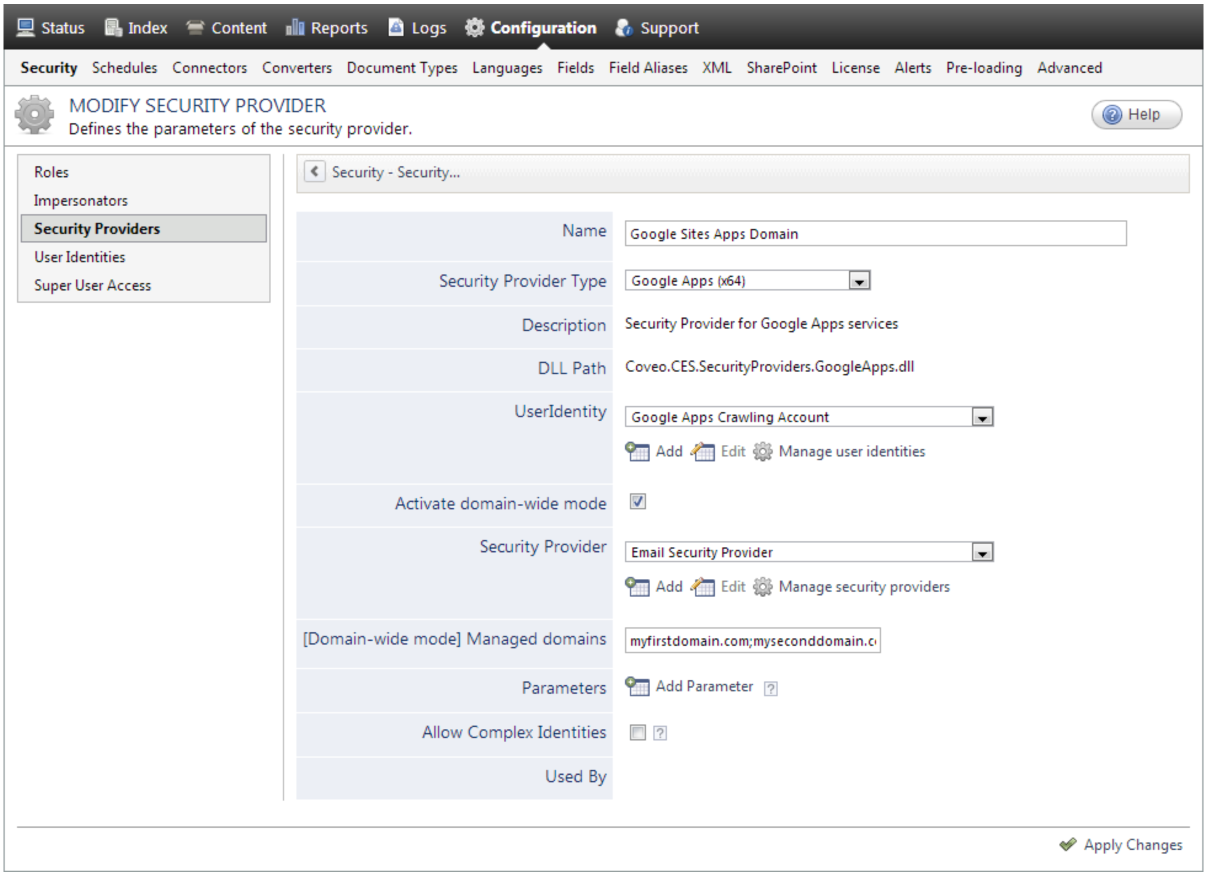
The Coveo Google Sites connector supports the Google Sites site-level permissions.

When you want users searching for Google Sites content in a Coveo search interface to only see the content to which they have access in Google Sites, the connector needs a security provider to be able to index the permissions for each indexed Google Sites item.

Note: You can get familiar with how Coveo components deal with permissions on documents both at indexing and query time.

To configure a Google Drive for Work security provider

1. On the Coveo server, access the Administration Tool.
2. Select **Configuration** > **Security**.
3. In the navigation panel on the left, click **Security Providers**.
4. In the **Security Providers** page, click **Add** to create a new security provider.
5. In the **Modify Security Provider** page:



MODIFY SECURITY PROVIDER

Defines the parameters of the security provider.

Name	Google Sites Apps Domain
Security Provider Type	Google Apps (x64)
Description	Security Provider for Google Apps services
DLL Path	Coveo.CES.SecurityProviders.GoogleApps.dll
UserIdentity	Google Apps Crawling Account
Activate domain-wide mode	☑
Security Provider	Email Security Provider
[Domain-wide mode] Managed domains	myfirstdomain.com;myseconddomain.c
Parameters	
Allow Complex Identities	☐
Used By	

a. Configure the following required parameters:

Name

Choose a meaningful name to identify the security provider.

Example: When the security provider is to be used by a source indexing Google Sites associated with an Apps domain, you can enter `Google Sites Apps Domain Security Provider`

Security Provider Type

In the drop-down list, select **Google Apps (x64)**.

User Identity

In the drop-down list, select the user identity that you selected or created previously (see Google Sites Connector Deployment Overview).

Activate domain-wide mode

You must select this option when you plan to use this security provider with a **Google Sites** source that indexes a Google Sites for an Apps domain.

Clear this option when the security provider is for a source indexing Google Sites of an individual user.

Security Provider

Select the security provider that you selected or created to allow this security provider to resolve and expand the groups (see Google Sites Connector Deployment Overview).

[Domain-wide mode] Managed domains

Enter the domain that you want to index. When your Google Apps account contains more than one domain, you can enter a semicolon-separated list of domains to index. The security provider will resolve and expand groups for the specified domain(s).

Examples:

- One domain: `mydomain.com`
- Multiple domains: `myfirstdomain.com;my.second.domain.com`

Important: The domain(s) specified in this list must match the one(s) specified in the **Addresses** box of the source that will use this security provider (see Configuring and Indexing a Google Sites Source).

Allow Complex Identities

Leave this option cleared as it does not apply to Google Sites.

b. Click **Apply Changes**.

What's Next?

Create and index a source (see Configuring and Indexing a Google Sites Source).

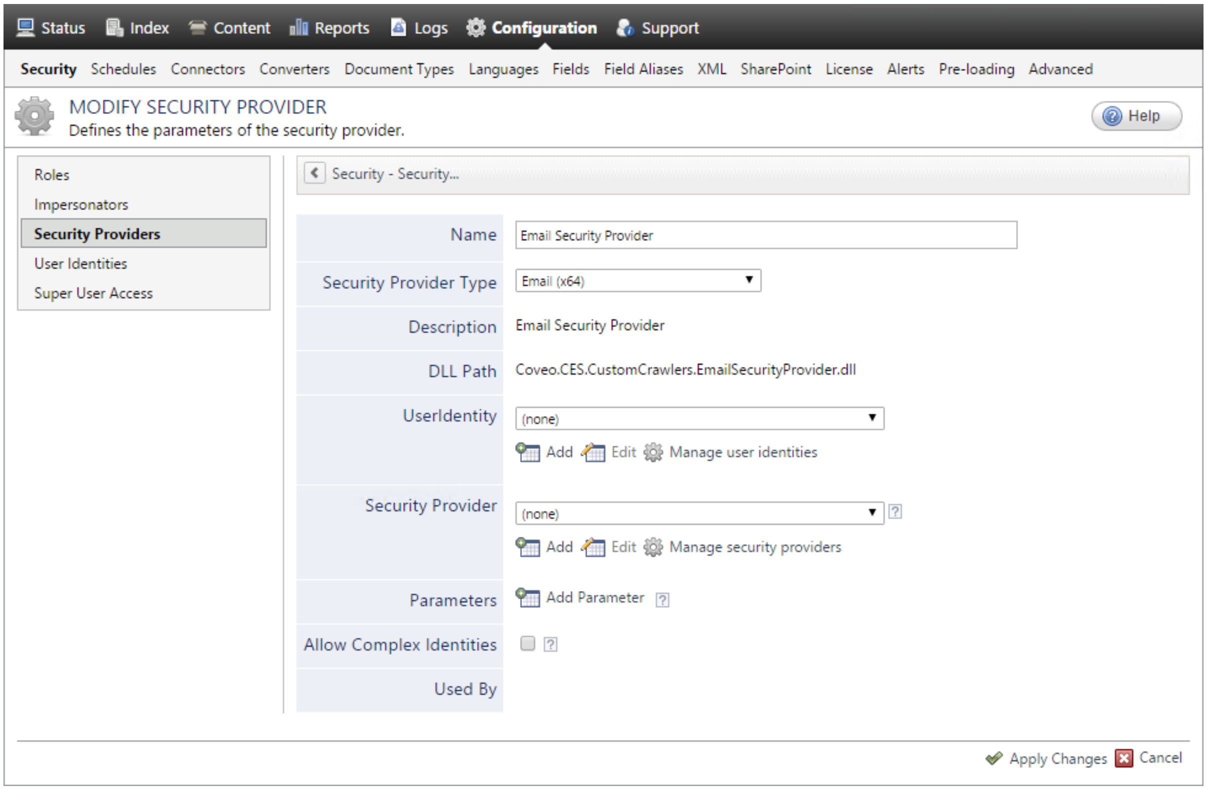
6.1 Configuring an Email Security Provider

An Email security provider is a simple email user identity container that can be used by another security provider to recognize users by their email addresses. When used by more than one security providers attached to sources of various types, an email security provider can act as a single sign-on system. An Email security provider does not connect to any system so it does not need a user identity.

Note: You can get familiar with how Coveo components deal with permissions on documents both at indexing and query time.

To configure an Email security provider

1. On the Coveo server, access the Administration Tool.
2. On the menu, select **Configuration** > **Security**.
3. In the navigation panel on the left, select **Security Providers**.
4. In the **Security - Security Providers** page, click **Add**.
5. In the **Modify Security Provider** page:



MODIFY SECURITY PROVIDER
Defines the parameters of the security provider.

Name: Email Security Provider
Security Provider Type: Email (x64)
Description: Email Security Provider
DLL Path: Coveo.CES.CustomCrawlers.EmailSecurityProvider.dll
UserIdentity: (none)
Security Provider: (none)
Parameters
Allow Complex Identities: □
Used By

   a. In the **Name** box, enter a name of your choice for your Email security provider.
   b. In the **Security Provider Type** list, select **Email**.

> **Note:** CES 7.0.5785 to 7.0.5935 (August to September 2013) The Email security provider DLL file is missing in the CES distribution so you will not see the **Email** option in the **Security Provider Type** list.
>
> To resolve this issue:
>
> i. Contact Coveo Support to get a copy of the `Coveo.CES.CustomCrawlers.EmailSecurityProvider.dll` file.
>
> ii. When you receive the file, using an administrator account, connect to the Coveo Master server, and then copy the file to the `[CES_Path]\bin` folder.
>
> iii. When your Coveo instance includes a Mirror server, also copy the file to the `[CES_Path]\bin` folder on the Coveo Mirror server.
>
> iv. Restart the CES service so that the new DLL is recognized.

c. In the **User Identity** list, leave **(none)**.

d. CES 7.0.7814+ (August 2015) (Optional) In the **Security Provider** list, select another security provider to map Email identities to another identity type.

> **Example:** You want to map Email identities to Active Directory (AD) ones so you select an LDAP Lookup security provider that is chained to an AD security provider. The LDAP Lookup security provider is then able to find a user in AD from his email and extracts his User Principal Name (UPN), thus allowing a mapping of the Email identity to an AD one. Contact Coveo Support for assistance on how to create an LDAP Lookup security provider.

e. Leave the **Allow Complex Identities** option cleared as it does not apply to this type of security provider.

f. Click **Apply Changes**.

What's Next?

Configure a security provider that will use this Email security provider.

6.2 Configuring an Active Directory Security Provider

You must use an Active Directory (AD) security provider when you create a source to index the content of an Active Directory domain. Other security providers may need to use an Active Directory security provider to expand, map, or resolve users or groups defined in Active Directory.

Coveo Enterprise Search (CES) comes with a default **Active Directory** security provider to which no user identity is assigned. In this case, the **Active Directory** security provider takes the CES service account as the user to access AD. When CES is in the same domain as AD, you can use the default **Active Directory** security provider as is. No configuration is needed.

You may need to create another Active Directory security provider only when CES and AD are in different and untrusted domains. In this case, you only need to assign a user identity containing any user that has access to the other domain to be able to use the security provider to expand, map, or resolve users or groups defined in Active Directory of this domain.

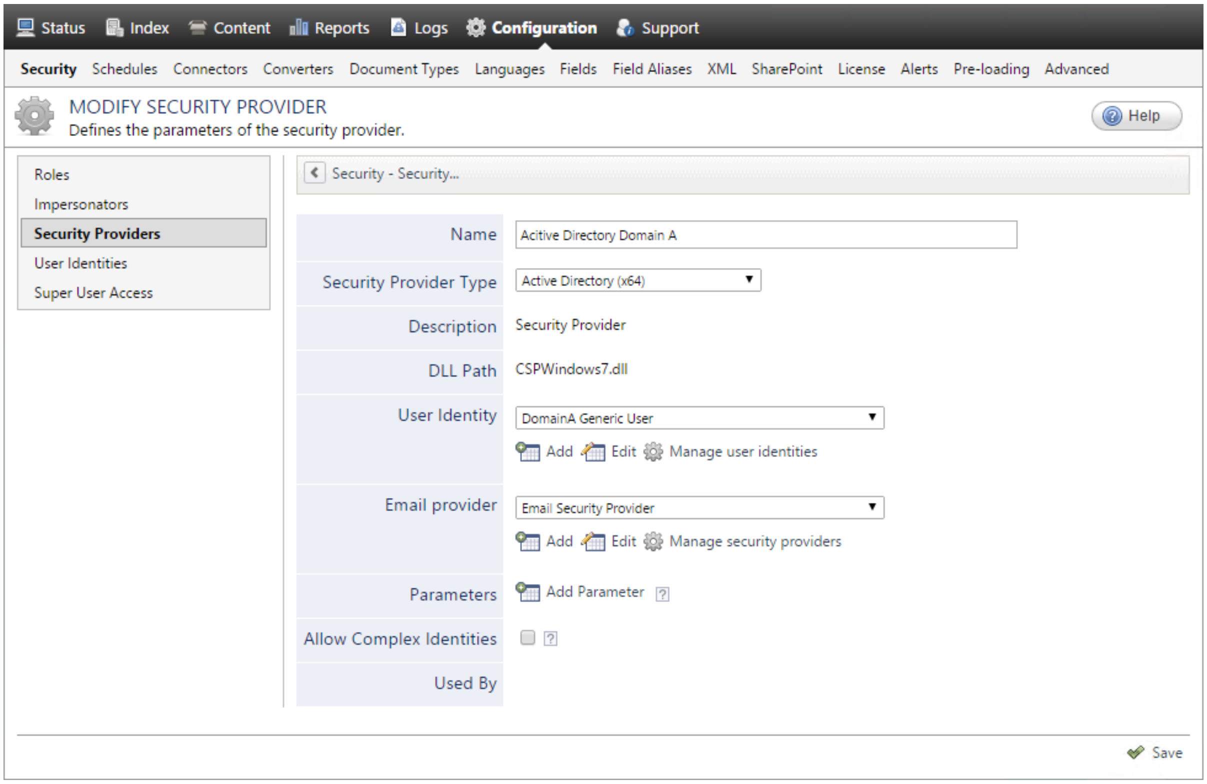
Note: You can get familiar with how Coveo components deal with permissions on documents both at indexing and query time.

To create or modify an Active Directory security provider

1. On the Coveo server, access the Administration Tool.
2. Select **Configuration** > **Security**.
3. In the navigation panel on the left, select **Security Providers**.
4. In the **Security Providers** page:
   - Click **Add** to create a new security provider.

     OR

   - Click an existing Active Directory security provider to modify it.
5. In the **Modify Security Provider** page:



Name	Active Directory Domain A
Security Provider Type	Active Directory (x64)
Description	Security Provider
DLL Path	CSPWindows7.dll
User Identity	DomainA Generic User
Email provider	Email Security Provider
Parameters	
Allow Complex Identities	☐
Used By	

   a. In the **Name** box, enter a name to identify this security provider.

   b. In the **Security Provider Type** drop-down list:

   i. On a 32-bit server, select **Active Directory (x86)**.

   ii. On a 64-bit server, select **Active Directory (x64)**.

c. In the **User Identity** section:

   i. In the drop-down list, select a user identity containing an account that has access to the desired domain.

   > **Example:** When the user identity contains the `domainA\OneUsername` account, the security provider connects to *Domain A* Active Directory.

   > **Note:** When **User Identity** is set to **(none)**, the security provider takes the CES service account by default.

   ii. When needed, click **Add**, **Edit**, or **Manage user identities** respectively to create, modify, or manage user identities.

d. CES 7.0.7338+ (January 2015) In the **Email Provider** section:

   i. In the drop-down list, select the email provider that recognizes your users by their email addresses.

   > **Note:** When you do not want to map Active Directory (AD) users to their email, select **(none)**.

   ii. When needed, click **Add**, **Edit**, or **Manage security providers** respectively to create, modify, or manage email security providers.

e. In the **Parameters** section, in rare cases the Coveo Support could instruct you to click **Add Parameters** to specify other security provider parameter names and values that could help to troubleshoot security provider issues.

f. Leave the **Allow Complex Identities** option cleared as it does not apply to this type of security provider.

g. Click **Save** or **Apply Changes**, depending whether you are creating or modifying a security provider.

What's Next?

When you are creating or modifying the security provider:

- For an Active Directory source, configure and index the source.
- To be used by another security provider, create or modify the other security provider.

7. Creating a Custom Google Sites Connector Mapping File

The Google Sites connector comes with a default mapping file (CES 7.0.7183+ (November 2014)) which allows to index the retrieved Google Sites default metadata. However, when you want to modify the default behavior of the Google Sites connector, creating a custom mapping file is required.

The following table presents Google Sites metadata retrieved by the connector. Metadata prefixed with `coveo_` are extra metadata created by the Coveo connector.

Note: CES 7.0.7183+ (November 2014) Metadata prefixed with `coveo_` and other absent metadata from the default mapping file are contained in the default field set (`[CES_Path]\Bin\Coveo.CES.CustomCrawlers.GoogleSites.FieldSet.xml`).

Name	Type	Description
`coveo_id`	`String`	Item unique identifier
`coveo_parent_id`	`String`	Parent item unique identifier
`coveo_domain_name`	`String`	The domain name of the Google Site that contains this item.
`coveo_site_name`	`String`	The site name of the Google Site that contains this item.
`coveo_item_type`	`String`	Item type formatted ready to be displayed.
`title`	`String`	Item title
`alternate_uri`	`String`	Item alternate URI (clickable link to the page or item in Google Sites)
`clickable_link`	`String`	Item clickable link
`is_draft`	`Boolean`	Indicates if the item is in draft state or not.
`kind`	`String`	Item kind (item type)
`folder`	`String`	Item folder name (if any)
`categories`	`String`	Item categories (tags)
`summary`	`String`	Item summary
`edited`	`DateTime`	Item edited date and time
`published`	`DateTime`	Item published date and time
`updated`	`DateTime`	Item updated date and time
`author_names`	`String`	Item author names
`author_emails`	`String`	Item author emails

Name	Type	Description
`content_source`	`String`	Item content source (link pointing to the attachment or web attachment)
`page_name`	`String`	Item page name in the URL
`revision`	`Integer`	Item current revision number

Note: CES 7.0.7104– (October 2014) When you want to index Google Sites metadata, you must create a custom mapping file that will be used by your Google Sites source. The mapping file specifies which Google Sites metadata will be mapped to which Coveo index custom fields.

Below is a sample Google Sites mapping file with useful mappings for Google Sites metadata.

```
<?xml version="1.0" encoding="utf-8" ?>
<Mappings>
 <Version>1</Version>
 <CommonMapping>
   <Fields>
     <!-- Google Sites custom fields -->
     <Field name="sitedomainname">%[coveo_domain_name]</Field>
     <Field name="siteitemtype">%[coveo_item_type]</Field>
     <Field name="sitename">%[coveo_site_name]</Field>
     <Field name="siteauthoremails">%[author_emails]</Field>
     <Field name="sitekind">%[kind]</Field>
     <Field name="sitecategories">%[categories]</Field>
   </Fields>
 </CommonMapping>
 <Mapping type="attachment">
   <Fields>
     <!-- Google Sites custom fields -->
     <Field name="sitefoldername">%[folder]</Field>
   </Fields>
 </Mapping>
 </Mapping>
 <Mapping type="listitem">
   <Title>List item added by %[author_names] at %[updated]</Title>
   <Fields>
     <Field name="sysauthor">%[author_names]</Field>
   </Fields>
 </Mapping>
 <Mapping type="page">
   <Fields>
     <Field name="sysauthor">%[author_names]</Field>
   </Fields>
 </Mapping>
 <Mapping type="website">
   <Body><![CDATA[ <html>%[title] <p>%[summary]</p></html> ]]></Body>
   <Fields>
     <Field name="sysauthor">%[author_names]</Field>
   </Fields>
 </Mapping>
 <Mapping type="webattachment">
   <Fields>
     <!-- Google Sites custom fields -->
     <Field name="sitefoldername">%[folder]</Field>
   </Fields>
 </Mapping>
</Mappings>
```

To create a custom Google Sites connector mapping file

1. Copy the content of the sample mapping file.
2. Using a text editor:
   a. Edit the file to include the desired Google Site metadata mapping to custom index fields.
   b. Ensure that your custom mapping file respects the standard mapping file format.
   c. Save the file.
3. Copy your custom mapping file on the Coveo Master server in a folder accessible to CES.

   Example: Copy the mapping file to `D:\CES7\Config\MyGoogleSitesMappingFile.xml`

What's Next?

Consider using the sample Google Sites field set (see Google Sites Connector Deployment Overview).

Ensure that all the Coveo fields included in your mapping file are created in the index.

8. Configuring and Indexing a Google Sites Source

A source defines a set of configuration parameters for one or more Google Sites for one Google Apps account.

Note: Create separate Google Sites sources when:

- You have more than one Google apps account to manage your Google Sites domains.
- One source Google Sites associated with a private Google account.

To configure and index a Google Sites source

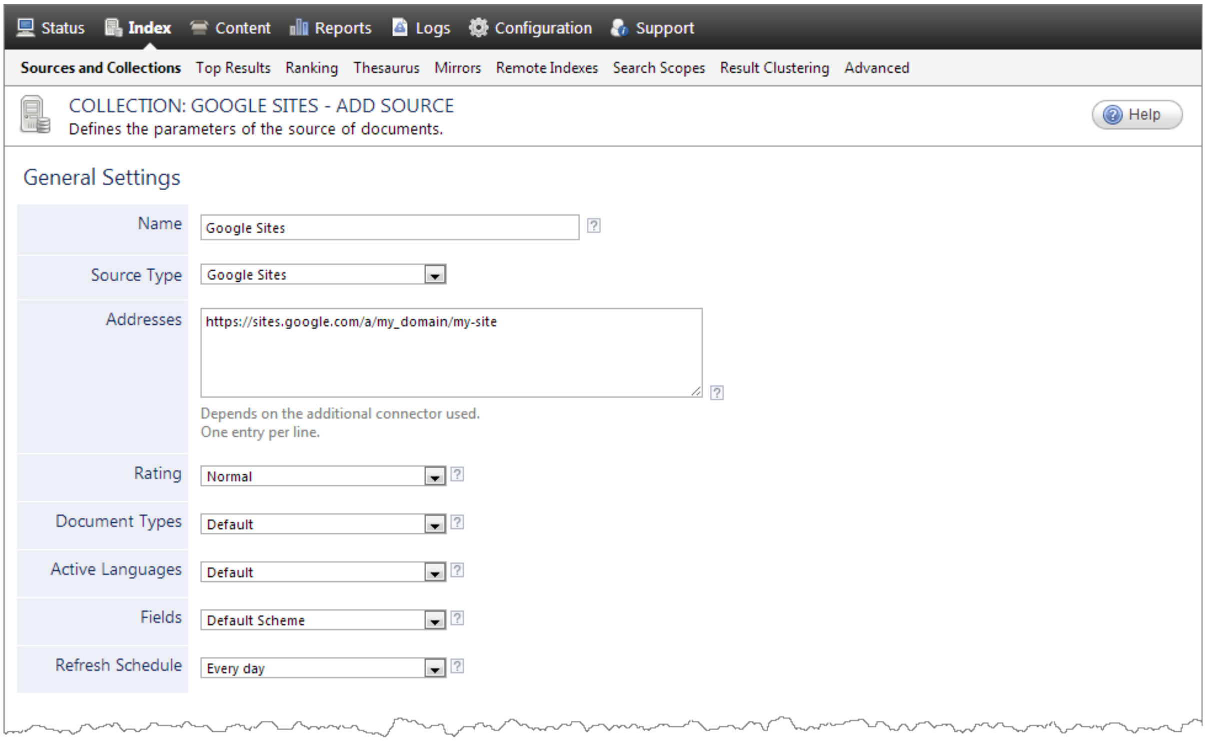
1. On the Coveo server, access the Administration Tool.
2. Select **Index** > **Sources and Collections**.
3. In the **Collections** section:

   a. Select an existing collection in which you want to add the new source.

   OR

   b. Click **Add** to create a new collection.
4. In the **Sources** section, click **Add**.

   The **Add Source** page that appears is organized in three sections.
5. In the **General Settings** section of the **Add Source** page:



COLLECTION: GOOGLE SITES - ADD SOURCE

Defines the parameters of the source of documents.

General Settings

Name	Google Sites
Source Type	Google Sites
Addresses	https://sites.google.com/a/my_domain/my-site Depends on the additional connector used. One entry per line.
Rating	Normal
Document Types	Default
Active Languages	Default
Fields	Default Scheme
Refresh Schedule	Every day

a. Enter the appropriate value for the following required parameters:

Name

Enter a descriptive name of your choice for the connector source.

Example: `Google Sites`

Source Type

Select the connector used by this source. In this case, select **Google Sites**.

Note: If you do not see **Google Sites**, your environment does not meet the requirements (see "Google Sites Connector Requirements" on page 6).

Addresses

Enter the address of one or more specific Google Sites in one of the following formats:

- For a private account: `https://sites.google.com/site/<my_site>`
- For a domain account: `https://sites.google.com/a/<my_domain>/<my_site>`

OR

Enter a starting address to use auto-discovery to crawl all sites accessible to the connector using one of the following formats:

- For a private account: `https://sites.google.com/site`
- For a domain account: `https://sites.google.com/a/<my_domain>`

Notes:

- The Google Sites returned by the auto-discovery feature are the ones to which the connector was granted access to when the OAuth2 refresh token was generated.
- Only auto-discovery of all accessible sites within a single domain is supported, not for all domains of a specific Google Apps account.
- The auto-discovery only returns sites with sharing permissions explicitly allowing the crawling user; it does not return sites allowing everyone from the domain nor does it consider administrator permissions which grants access to all web sites of the Google App domain.

Fields

Select the field set that you created earlier (see Google Sites Connector Deployment Overview).

Refresh Schedule

Time interval at which the index is automatically refreshed to keep the index content up-to-date. By default, the **Every day** option instructs CES to refresh the source everyday at 12 AM. Because the incremental refresh takes care of maintaining the source up-to-date, you can select a longer interval such as **Every Sunday**.

b. Review the value for the following parameters that often do not need to be modified:

Rating

Change this value only when you want to globally change the rating associated with all items in this source relative to the rating to other sources.

> **Example:** When a source replaces a legacy system, you may want to set this parameter to **High**, so that in the search interface, results from this source appear earlier in the list compared to those from legacy system sources.

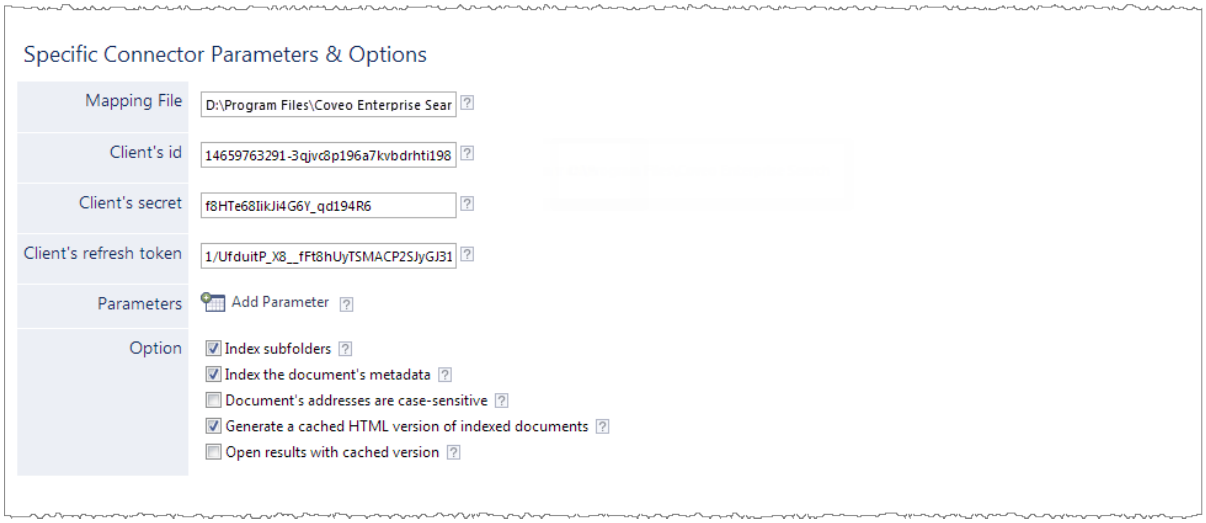
Document Types

If you defined a custom document type set for this source, select it.

Active Languages

If you defined custom active language sets, ensure to select the most appropriate for this source.

6. In the **Specific Connector Parameters & Options** section of the **Add Source** page:



Specific Connector Parameters & Options

Mapping File	D:\Program Files\Coveo Enterprise Sear
Client's id	14659763291-3qjvc8p196a7kvbdrhti198
Client's secret	f8HTe68IikJi4G6Y_qd194R6
Client's refresh token	1/UfduitP_X8__fFt8hUyTSMACP2SJyGJ31
Parameters	Add Parameter
Option	☑ Index subfolders ☑ Index the document's metadata ☐ Document's addresses are case-sensitive ☑ Generate a cached HTML version of indexed documents ☐ Open results with cached version

a. In the **Mapping File** box, the path to the default mapping file that defines how the connector handles metadata often does not need to be changed.

Notes:

- CES 7.0.7256– (December 2014) Enter the path to the default mapping file that defines how the connector handles metadata. You can leave this box empty, in which case no Google Sites metadata will be indexed.

  Example: `D:\Program Files\Coveo Enterprise Search 7\Bin\Coveo.CES.CustomCrawlers.GoogleSites.MappingFile.xml`

- CES 7.0.7104– (October 2014) If you create a custom mapping file, enter the path where you saved your file on the Coveo server (see "Creating a Custom Google Sites Connector Mapping File" on page 18). You can leave this box empty, in which case no Google Sites metadata will be indexed.

  Example: `D:\CES7\Config\MyGoogleSitesMappingFile.xml`

b. Using the following parameters, authorize the Coveo crawler to access the Google Sites:

Client's id

Enter the Client ID value that you got earlier (see "Getting Google Sites Client ID and Client Secret values" on page 7).

Client's secret

Enter the Client Secret value that you got earlier (see "Getting Google Sites Client ID and Client Secret values" on page 7).

Client's refresh token

Enter the OAuth2 refresh token value that you got earlier (see "Getting a Google Sites OAuth2 Refresh Token" on page 9).

c. Click **Add Parameter** when you want to show and change the value of advanced source parameters (see "Modifying Hidden Google Sites Source Parameters" on page 27).

d. The **Option** check boxes generally do not need to be changed:

Index Subfolders

This parameter is not taken into account for this connector.

Index the document's metadata

When selected, CES indexes all the document metadata, even metadata that are not associated with a field. The orphan metadata are added to the body of the document so that they can be searched using free text queries.

When cleared (default), only the values of system and custom fields that have the **Free Text Queries** attribute selected will be searchable without using a field query.

Example: A document has two metadata:

- `LastEditedBy` containing the value `Hector Smith`
- `Department` containing the value `RH`

In CES, the custom field `CorpDepartment` is bound to the metadata `Department` and its **Free Text Queries** attribute is selected.

When the **Index the document's metadata** option is cleared, searching for `RH` returns the document because a field is indexing this value. Searching for `hector` does not return the document because no field is indexing this value.

When the **Index the document's metadata** option is selected, searching for `hector` also returns the document because CES indexed orphan metadata.

Document's addresses are case-sensitive

Leave the check box cleared. This parameter needs to be checked only in rare cases for systems in which distinct documents may have the same name but different casing.

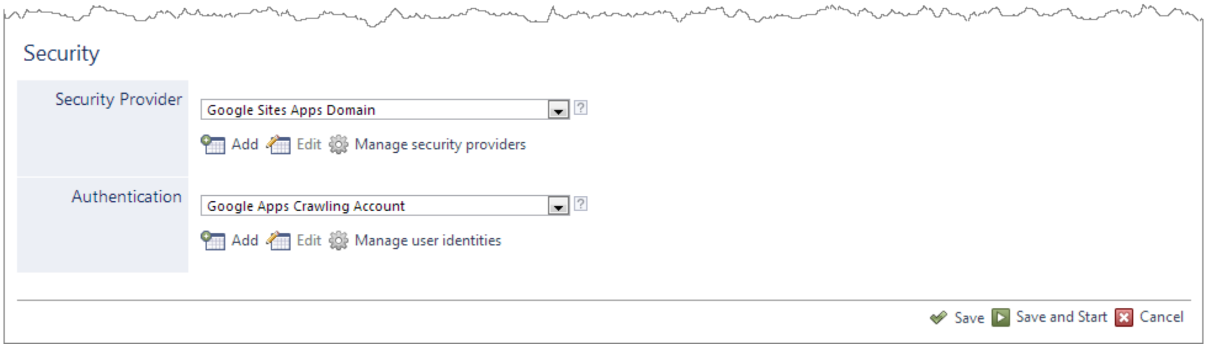
Generate a cached HTML version of indexed documents

When you select this check box (recommended), at indexing time, CES creates HTML versions of indexed documents. In the search interfaces, users can then more rapidly review the content by clicking the **Quick View** link rather than opening the original document with the original application. Consider clearing this check box only when you do not want to use **Quick View** links or to save resources when building the source.

Open results with cached version

Leave this check box cleared (recommended) so that in the search interfaces, the main search result link opens the original document with the original application. Consider selecting this check box only when you do not want users to be able to open the original document but only see the HTML version of the document as a **Quick View**. In this case, you must also select **Generate a cached HTML version of indexed documents**.

7. In the **Security** section of the **Add Source** page:



Security

Security Provider: Google Sites Apps Domain

Authentication: Google Apps Crawling Account

a. When you chose to index Google Sites permissions, in the **Security Provider** drop-down list, select the Google Sites security provider that you created for this source (see "Configuring a Google Sites Security Provider" on page 12).

b. In the **User Identity** drop-down list, select the user identity that you created for this source (see Google Sites Connector Deployment Overview).

8. Click **Save** to save the source configuration.

9. When you chose to not index Google Sites permissions, you can set source level permissions that apply to all documents in the source:

   a. In the navigation panel on the left, click **Permissions**.

   b. In the **Permissions** page, select **Specify the security permissions** to index.

   c. In the **Allowed Users** and **Denied Users** boxes, enter the users and groups that you respectively want to allow or deny to see search results from this source. The default is to allow **everyone (Active Directory Group)**.

   d. Click **Apply Changes**.

10. When you are ready to start indexing the Google Sites source, click **Rebuild**.

11. Validate that the source building process is executed without errors:

    - In the navigation panel on the left, click **Status**, and then validate that the indexing proceeds without errors.

      OR

    - Open the CES Console to monitor the source building activities.

What's Next?

Set an incremental refresh schedule for your source.

Consider modifying some hidden source parameters to try resolving issues (see "Modifying Hidden Google Sites Source Parameters" on page 27).

9. Modifying Hidden Google Sites Source Parameters

The **Add Source** and **Source: ... General** pages of the Administration Tool present the parameters with which you can configure the connector for most Google Sites setups. More advanced and more rarely used parameters are hidden. You can choose to make one or more of these parameters appear in the **Add Source** and **Source: ... General** pages of the Administration Tool so that you can change their default value. Consider changing values of hidden parameters when you encounter issues.

The following list describes the advanced hidden parameters available with Google Sites sources. The parameter type (integer, string, etc.) appears between parentheses following the parameter name.

ResultsPerPage (Integer)

Number of items to fetch per request made to Google Sites. The default value is `100`. The minimum value is `1`. A small value (not recommended) forces the connector to make small but frequent queries to Google Sites. A larger value (recommended) leads to larger and less frequent queries.

NumberOfRetries (Integer)

This parameter determines the number retries allowed when a recoverable call to Google Sites API fails. The default value is `2`.

To modify hidden Google Sites source parameters

1. Refer to "Adding an Explicit Connector Parameter" on page 28 to add one or more Google Sites source parameters.
2. For a new Google Sites source, access the **Add Source** page of the Administration Tool to modify the value of the newly added advanced parameter:
    a. Select **Index** > **Sources and Collections**.
    b. Under **Collections**, select the collection in which you want to add the source.
    c. Under **Sources**, click **Add**.
    d. In the **Add Source** page, edit the newly added advanced parameter value.
3. For an existing Google Sites source, access the **Source: ... General** page of the Administration Tool to modify the value of the newly added advanced parameter:
    a. Select **Index** > **Sources and Collections**.
    b. Under **Collections**, select the collection containing the source you want to modify.
    c. Under **Sources**, click the existing Google Sites source in which you want to modify the newly added advanced parameter.
    d. In the **Source: ... General** page, edit the newly added advanced parameter value.
4. Rebuild your Google Sites source to apply the changes to the parameters.

9.1 Adding an Explicit Connector Parameter

Connector parameters applying to all sources indexed using this connector are called explicit parameters.

When you create or configure a source, the Coveo Enterprise Search (CES) 7.0 Administration Tool presents parameters with which you can configure the connector for most setups. For many connectors, more advanced and more rarely used parameters also exist but are hidden by default. CES then uses the default value associated with each of these hidden parameters.

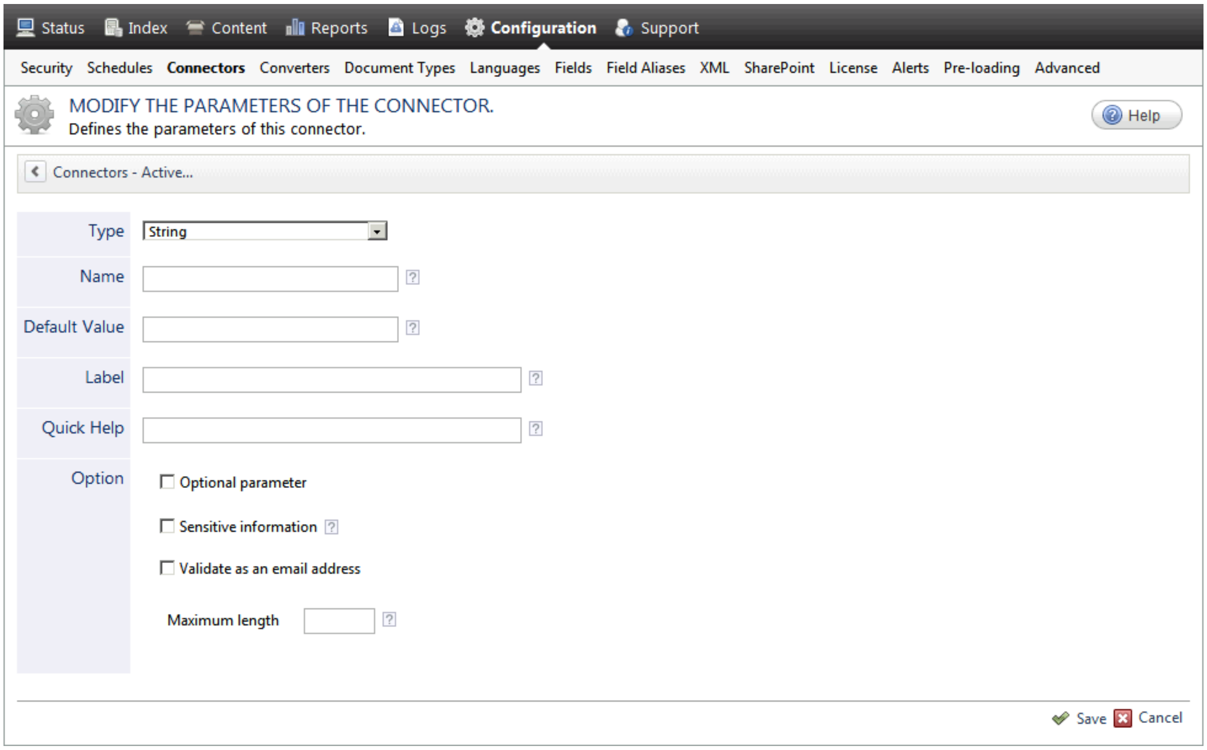
You can however choose to make one or more of these parameters appear in the **Add Source** and **Source: ... General** pages of the Administration Tool so that you can change their default value.

To add an explicit connector parameter

1. On the Coveo server, access the Administration Tool.
2. Select **Configuration** > **Connectors**.
3. In the list on the **Connectors** page, select the connector for which you want to show advanced hidden parameters.
4. In the **Parameters** section of the selected connector page, click **Add Parameter** for each hidden parameter that you want to modify.

   Note: The **Add Parameter** button is present only when hidden parameters are available for the selected connector.

5. In the **Modify the parameters of the connector** page:



MODIFY THE PARAMETERS OF THE CONNECTOR.
Defines the parameters of this connector.

Type	String
Name	
Default Value	
Label	
Quick Help	
Option	□ Optional parameter □ Sensitive information □ Validate as an email address Maximum length

a. In the **Type** list, select the parameter type as specified in the parameter description.

b. In the **Name** box, type the parameter name exactly as it appears in the parameter description. Parameter names are case sensitive.

c. In the **Default Value** box, enter the default value specified in the parameter description.

> **Important:** Do not set the value that you want to use for a specific source. The value that you enter here will be used for all sources defined using this connector so it must be set to the recommended default value. You will be able to change the value for each source later, in the **Add Source** and **Source: ... General** pages of the Administration Tool.

d. In the **Label** box, enter the label that you want to see for this parameter.

> **Example:** To easily link the label to the hidden parameter, you can simply use the parameter name, and if applicable, insert spaces between concatenated words. For the **BatchSize** hidden parameter, enter `Batch Size` for the label.

> **Note:** To create multilingual labels and quick help messages, use the following syntax: `<@ln>text</@>`, where *ln* is replaced by the language initials—the languages of the Administration Tool are English (en) and French (fr).

> **Example:** `<@fr>Chemin d'accès du fichier de configuration</@><@en>Configuration File Path</@>` is a label which is displayed differently in the French and English versions of the Administration Tool.

> **Tip:** The language of the Administration Tool can be modified by pressing the following key combination: `Ctrl+Alt+Page Up`.

e. Optionally, in **Quick Help**, enter the help text that you want to see for this parameter when clicking the question mark button ? that will appear beside the parameter value.

> **Tip:** Copy and paste key elements of the parameter description.

f. When **Predefined values** is selected in the **Type** parameter, in the **Value** box that appears, enter the parameter values that you want to see available in the drop-down parameter that will appear in the Administration Tool interface. Enter one value per line. The entered values must exactly match the values listed in the hidden parameter description.

g. Select the **Optional parameter** check box when you want to identify this parameter as an optional parameter. When cleared, CES does not allow you to save changes when the parameter is empty. This parameter does not appear for **Boolean** and **Predefined values** parameter types.

h. Select the **Sensitive information** check box for password or other sensitive parameter so that, in the Administration Tool pages where the parameter appears, the typed characters appear as dots to mask them. This parameter appears only for the **String** type.

Example: When you select the **Sensitive information** check box for a parameter, the characters typed appear as follows in the text box:

●●●●

i. Select the **Validate as an email address** check box when you want CES to validate that the text string that a user enters in this parameter respects the format of a valid email address. This parameter appears only for the **String** type.

j. In the **Maximum length** box, enter the maximum number of characters for the string. This parameter appears only for the **String** type. When you enter `0`, the length of the string is not limited.

k. Click **Save**.

6. Back in the **Connector** page, click **Apply Changes**.

   The hidden parameter now appears in the **Add Source** and **Source: ... General** pages of the Administration Tool for the selected source. You can change the parameter value from these pages. Refer to the documentation for each connector for details.

Note: When you want to modify a hidden source parameter, you must first delete it, and then redefine it with the modified values.