

Files: Google Drive, SharePoint and OneDrive, Amazon S3

From the SourceLace manual: How to connect file stores so people can find files where they already are and read the text inside them. Everything is read-only, and files are read only when someone asks, never copied or indexed.

SourceLace version 0.1.0 · Made October 4, 2026

The latest version is at <https://sourcelace.onrender.com/docs>

Files: Google Drive, SharePoint and OneDrive, Amazon S3

How to connect file stores so people can find files where they already are and read the text inside them. Everything is read-only, and files are read only when someone asks, never copied or indexed.

Status: in testing. Built against each service's documented API and covered by automated tests against simulated services; not yet used with a customer's live files.

How it works

- **Read on demand, never indexed.** SourceLace does not copy, crawl or index anyone's files. When someone asks to read a file, SourceLace downloads that one file, pulls the text out in memory, and returns it in chunks. The chunks are held in memory like any other result (up to 30 minutes) and then forgotten.
- **Finding files uses each store's own search:** Google Drive's search, Microsoft Search for SharePoint and OneDrive, and key names for S3 (S3 cannot search inside files).
- **What it can read:** PDF, Word (.docx), Excel (.xlsx), PowerPoint (.pptx), CSV, plain text (and Markdown, JSON, HTML), and Google Docs, Sheets and Slides. Older .doc, .xls and .ppt files, images, and scanned PDFs without a text layer are not read; SourceLace says so instead of guessing. Password-protected PDFs cannot be read.
- **Limits:** files up to 25 MB; up to 400,000 characters of text per file (longer files are marked `text_truncated`); PDFs up to 1,000 pages; Office files that unpack to more than 250 MB are refused.

Source	Kind	Who decides what someone can read
Google Drive	google_drive	Each person signs in with their own Google account. Drive's own sharing applies.
SharePoint and OneDrive	microsoft_files	Each person signs in with their own work or school account. SharePoint and OneDrive permissions apply.
Amazon S3	s3	The source, not the person. See Amazon S3 .

How people use it

Every file source takes one query language, `files`: a small JSON object with an action.

Action	Example	What it does
search	<pre>{"action": "search", "text": "corvanta renewal"}</pre>	Finds files with the store's own search. "match": "name" matches names only; the default "content" also matches text inside files where the store supports it. Optional: "folder", "modified_after": "2026-09-01", "limit" (up to 200).

Action	Example	What it does
list	<code>{"action": "list", "folder": "root"}</code>	The files and folders in one folder. root is My Drive, your OneDrive, or the top of the bucket.
drives	<code>{"action": "drives"}</code>	Where files live: shared drives, SharePoint document libraries (add "text": "sales" to find sites by name), or the source's buckets.
read	<code>{"action": "read", "file": "<id>"}</code>	The file's text, in chunks of 4,000 characters ("chunk_chars" up to 20,000), 10 chunks at a time ("chunks" up to 50). "from_chunk": 10 reads further.

get_record with object files and a file id returns the file's details and the first 20,000 characters of its text.

Google Drive

Kind: **Google Drive** (google_drive). No options.

Each person signs in with their own Google account through SourceLace's Google app (the same app as Gmail and Calendar), and is asked only for <https://www.googleapis.com/auth/drive.readonly> ("See and download all your Google Drive files"). Connecting Drive never gives SourceLace someone's mail, and connecting mail never gives it their files.

Your Google Workspace admin may need to allow SourceLace's app under **Google Admin console** → **Security** → **Access and data control** → **API controls** → **App access control**, if your organization restricts third-party apps. Google reviews apps that ask for Drive access; if Google shows an "unverified app" warning or refuses, contact support@sourcelace.com.

SharePoint and OneDrive

Kind: **SharePoint and OneDrive** (microsoft_files). No options.

One source covers both: each person's OneDrive (`{"action": "list"}`) and every SharePoint document library they can open (`{"action": "drives", "text": "sales"}`). Search covers both too. File ids look like b!abc...~01XYZ...; use them exactly as returned.

Each person signs in with their own work or school account through SourceLace's Microsoft app, and approves these **delegated** Microsoft Graph permissions (SourceLace acts as the person, with only their access): offline_access, User.Read, Files.Read.All, Sites.Read.All. If your organization requires an admin to approve apps, a Microsoft admin grants consent in the **Microsoft Entra admin center** → **Enterprise applications**.

Amazon S3

Kind: **Amazon S3** (s3).

S3 has no per-person sign-in, so **S3 access is per source, not per person**. Each S3 source reads with one set of AWS credentials, and everyone who can use that source reads with that same access. The IAM policy behind those credentials decides which buckets and files the source

can read. To narrow **who in your organization** can use an S3 source, use [access by group](#). Until you do, everyone in your organization can use every S3 source you add, so only add buckets that everyone may read.

SourceLace makes only two kinds of S3 call, both reads: **ListObjectsV2** (list and search) and **GetObject** (read). With a role, it also calls STS **AssumeRole** for short-lived credentials. It signs requests itself (AWS Signature Version 4) and uses no AWS SDK.

1. Create read-only access (your AWS admin)

AWS credentials are kept in SourceLace's server settings, never in a source's options (which admins can see), so **S3 is set up together with SourceLace support**. Create a read-only IAM policy, such as (change corvanta-docs and the reports/ folder to yours; drop the Condition to allow the whole bucket):

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "s3:ListBucket",
      "Resource": "arn:aws:s3:::corvanta-docs",
      "Condition": {
        "StringLike": {
          "s3:prefix": ["reports/*"]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": "s3:GetObject",
      "Resource": "arn:aws:s3:::corvanta-docs/report
s/*"
    }
  ]
}
```

Then, with support, choose one of:

- **A role SourceLace assumes:** put the policy on an IAM role in your account whose trust policy lets the AWS identity that support names assume it, with an **external id**. You give the source the role's ARN and external id, so no long-lived keys of yours leave your account.
- **Access keys:** an IAM user with the policy and an access key ("Application running outside AWS"), sent to support over a secure channel we agree with you. Support stores them under a name (such as default) in SourceLace's server settings.

2. Add the source (SourceLace admin)

Option	Type	Default	Example	What it is
buckets	List, separated by commas	(required)	corvanta-docs	The buckets this source reads.
region	Text	(required)	us-east-1	The buckets' AWS region.
prefix	Text	(none)	reports/	Only read under this folder. SourceLace enforces it too, before any call.
keys	Text	default	finance	The name of the credentials in SourceLace's server settings, as support tells you.
role_arn	Text	(none)	arn:aws:iam::123456789012:role/sourcelace-read	An IAM role to assume.
external_id	Text	(none)	corvanta-7f3a	The external id the role's trust policy expects.

People still click **Connect** once (there is no sign-in page; it only records that they connected).
list_sources shows the source's access as the key name or the role, not a person.

S3 search matches file names only, and looks through up to 10,000 keys per search: use prefix or folder to narrow big buckets.

When something goes wrong

What you see	What to do
"Google did not grant read access to Drive. Connect again and tick the box on Google's consent screen."	Connect again and tick the Drive box.
"'...' is ... MB; SourceLace reads files up to 25 MB."	The file is too large to read.
"The PDF is password-protected, so SourceLace cannot read it."	Remove the password, or read it another way.
"SourceLace could not read the text in '...': the file looks damaged or is in an unexpected format"	The file could not be read; open it in its own app to check.
"'...' is a folder. List it with ..."	Use list on a folder, read on a file.
"'...' is not a SharePoint or OneDrive id. Use an id from a search, list or drives result."	Use file ids exactly as SourceLace returned them.
"... needs a valid "buckets" (bucket names, separated by commas) in its options." (or "region", "role_arn", ...)	Fix that option on Manage sources .
"... uses the S3 keys named '...', which are not in SOURCELACE_S3_KEYS on the server."	The keys option names credentials SourceLace does not have. Check the name with support.
"AWS refused to let SourceLace assume the source's role (HTTP ...). Check role_arn, external_id and the role's trust policy."	Fix the role's trust policy, role_arn or external_id.
"Amazon S3 refused access (AccessDenied). The source's keys or role cannot read this; an admin can change its IAM policy."	Widen the IAM policy if the source should read it.
"This source only reads under ...".	The file is outside the source's prefix.
"'...' is not in this source's buckets (...)."	The folder is in a bucket the source does not list.
"Microsoft sign-in failed: ..."	Microsoft's own reason follows. If it mentions admin approval or consent, ask your Microsoft admin.