LAREG CORP · ISO DOCUMENT COLLECTION

LaReg ISO Document Hub User Manual

A plain-English guide to collecting, browsing and using every public SPP and PJM stakeholder document since January 1, 2025. No technical background needed.

Portal: lareg-iso-document-hub.pages.dev

Covers: SPP (spp.org) · PJM (pjm.com) · documents.csv · folder tree

Prepared by: Poornima Ramakrishnan · Version 1.0 · November 2026

Contents

1. What this is, in plain English
2. A two-minute tour of the portal
3. Finding a document
4. Opening the delivered files
5. What each CSV column means
6. Running the collector yourself
7. Keeping the collection up to date
8. Checking the results
9. Adding a folder, page or new ISO
10. Troubleshooting & FAQ
11. Glossary
12. Getting help

1. What this is, in plain English

LaReg needs every public document that two US grid operators, **SPP** and **PJM**, have posted since January 1, 2025: meeting agendas, minutes, presentations, studies, filings, tariffs and more. They are spread across hundreds of web folders and pages.

The **LaReg ISO Document Collector** is a small program that visits exactly the folders and pages listed in the brief, politely and slowly. It downloads each document once and files it in a tidy folder structure. It also writes one spreadsheet, `documents.csv`, describing every file: where it came from, its title, its dates and a fingerprint that proves it's intact.

Run it again next week and it only picks up what's new or changed. Nothing is duplicated and nothing is overwritten.

This portal is a window into a real sample run. You can browse, search and download the actual files the collector gathered, before the full collection is done.

2. A two-minute tour of the portal

The menu at the top has eight pages. Here's what each is for:

Overview

Headline numbers from the sample run, how the collector works, and what's been checked.

Document explorer

Search and filter every collected file. Click a row for its details and the original link.

Coverage

Every SPP folder and PJM page from the brief, what was found, and anything flagged.

Collector

How the collector works, its commands, and the polite-crawling rules.

Data model

The 15 CSV columns and the database design behind them.

Integrations

Copy-paste recipes for Snowflake, Google Sheets, Python and AI tools.

Downloads

The sample CSV, real files, reports and this manual as a PDF.

Milestones

The two-step delivery plan and what comes next.

The **sun / moon** button switches between light and dark mode, and your choice is remembered. The small **Feedback** tab on the side lets you send a note at any time.

3. Finding a document

- Open Document explorer.
- Type in the search box, for example *tariff*, *minutes* or a committee name. Press `/` from anywhere on the page to jump to it.
- Narrow it down with the buttons: **SPP** or **PJM**, a folder or committee, a file type, or a date range.
- Click any row to see full details: the title as shown on the site, posting and meeting dates, size, fingerprint and a link to the original.
- Click **Export CSV** to download just the rows you filtered.

The address bar keeps your filters, so you can bookmark or share a search.

4. Opening the delivered files

The full collection arrives in a shared Google Drive or Dropbox folder. It looks like this:

```
LaReg_ISO_Document_Collection/
├── documents.csv          ← the spreadsheet of every file
├── SPP/                   ← SPP folders, same names as on spp.org
├── PJM/                   ← PJM committees, by year and meeting date
└── _reports/              ← coverage, misses and check results
```

- **To browse:** open the SPP or PJM folder and click through it like any other folder. The names match the websites.
- **To see everything at once:** open `documents.csv` in Excel (Data → From Text/CSV → UTF-8) or Google Sheets (File → Import).
- **To find a file from the spreadsheet:** the `local_path` column is its exact location inside the folder.

Folders ending in `__unzipped` hold the contents of a zip that sits right next to it. The original zip is always kept too.

5. What each CSV column means

COLUMN	WHAT IT TELLS YOU
`source_iso`	SPP or PJM
`source_key`	Unique key: SPP doc_id, PJM full /-/media/... path (zip members: key!/member)
`source_url`	Full URL the file was downloaded from
`file_name`	The file name as saved
`file_extension`	pdf, docx, xlsx, zip, ...
`file_size_bytes`	Size of the downloaded file
`content_sha256`	SHA-256 of the file contents
`title`	Link text / title exactly as shown on the listing
`posting_date`	Posting date from the listing, YYYY-MM-DD
`container_id`	SPP folder ID or PJM committee code
`container_path`	SPP folder breadcrumb or PJM page path
`meeting_date`	Meeting date from the path or title, YYYY-MM-DD
`parent_zip`	File name of the parent .zip for extracted members, else blank
`local_path`	Relative path of the saved file in the delivered tree
`crawl_timestamp`	When the file was downloaded (UTC, ISO 8601)

6. Running the collector yourself

You don't need to do this, since the full collection is delivered to you. But if your team wants to run it, it takes about ten minutes to set up on any Windows, Mac or Linux computer.

1. **Install Python 3.11 or newer** from python.org. On Windows, tick *"Add Python to PATH"* during setup.

1. **Unzip the collector** into a folder, for example `C:\LaReg\collector`.
2. **Open a terminal in that folder** (Windows: right-click the folder → *Open in Terminal*) and install it once:

   ```
   python -m pip install -e .
   ```

3. **Start the full collection**:

   ```
   lareg-iso --output D:\LaReg_ISO_Document_Collection backfill --since 2025-01-01
   ```

   Leave it running. You'll see progress lines as it goes. A full run takes many hours because it is deliberately polite.
4. **Try a small sample first** if you like. This collects one SPP folder and one PJM committee:

   ```
   lareg-iso --output D:\LaReg_Sample backfill --spp-folders 31491 --pjm-pages committees/mrc
   ```

Disk space: SPP meeting packages can be 100–360 MB each, so allow at least 100 GB free for the full collection.

7. Keeping the collection up to date

After the first full run, use **incremental** mode. It only downloads files that are new or changed:

```
lareg-iso --output D:\LaReg_ISO_Document_Collection incremental
```

Running it twice in a row is safe. The second run downloads nothing and changes nothing.

To run it automatically every night on Windows:

1. Open *Task Scheduler* → *Create Basic Task* → name it *LaReg ISO incremental*.
2. Trigger: *Daily*, e.g. 2:00 AM.
3. Action: *Start a program*. Program: `lareg-iso`. Arguments: `--output D:\LaReg_ISO_Document_Collection incremental`.

On Mac or Linux, one cron line does the same: `0 2 * * * lareg-iso --output ~/LaReg incremental`.

8. Checking the results

Two built-in commands prove the collection is complete and correct:

```
lareg-iso --output D:\LaReg_ISO_Document_Collection verify
lareg-iso --output D:\LaReg_ISO_Document_Collection spot-check -n 10
```

- **verify** confirms that every row's file exists, that its fingerprint matches, that every zip was unpacked, and that counts and sizes add up for each site.
- **spot-check** picks 10 random rows per site and re-checks their title, date and link against the live website.
- **status** shows the last few runs and anything that still needs a retry.

Each check prints a clear PASS or FAIL line and saves a report in `_reports/`.

9. Adding a folder, page or new ISO

Everything the collector visits is listed in one settings file, `config/collector.yaml`. Open it in Notepad or any text editor:

- **New SPP folder:** add its number (the `id=` in its spp.org address) under `folders:`.
- **New PJM committee:** add its short code (the last part of its pjm.com address) under the right group in `pages:`.
- **Different start date:** change `since:`, or pass `--since 2024-01-01` on the command line.

Then run `incremental`, and only the new additions are collected. Adding a completely new grid operator (MISO, ERCOT, ...) needs a small code module, which the design already allows for.

10. Troubleshooting & FAQ

A file is listed on the site but isn't in my CSV.

Check its posting date first. Only files posted on or after the cutoff (Jan 1, 2025 by default) are collected. If the date is in range, look in _reports/misses.csv. Anything that failed is listed there with the reason, and the next run retries it automatically.

The run stopped halfway (power cut, closed laptop).

Just run the same command again. Files that were already saved are skipped, and half-downloaded files are thrown away and fetched again.

Why is the run so slow?

On purpose. The collector waits about 2 seconds between requests to each site so it never strains SPP's or PJM's servers. PJM pages also take ~15 seconds to answer. A full backfill takes many hours and that is expected.

A document changed on the website. Did I lose the old one?

No. The old file is kept and the new one is saved next to it as name.rev2.pdf, with its own CSV row. Nothing is ever overwritten.

What are the __unzipped folders?

When a download is a .zip, the collector keeps the zip and also unpacks it into a folder with the same name plus __unzipped. Each file inside gets its own CSV row, with parent_zip set to the zip's name.

Can I open the CSV in Excel without breaking it?

Yes. Use Data → From Text/CSV and choose UTF-8 so titles with special characters display correctly. Google Sheets reads it directly with File → Import.

Some PJM committees have nothing from early 2025.

Eight groups started meeting later or don't list 2025 materials. They're flagged on the Coverage page and in the README. That's the site, not a gap in the collector.

Does it need a special browser or login?

No. Both sites' documents are public and readable without a headless browser, login or API key.

11. Glossary

ISO / RTO

Independent System Operator / Regional Transmission Organization: the bodies that run the power grid and wholesale markets in a region.

SPP

Southwest Power Pool, the RTO for the central US. Its documents live in a numbered folder tree on spp.org.

PJM

PJM Interconnection, the RTO for 13 Mid-Atlantic and Midwest states. Its documents hang off committee and task-force pages on pjm.com.

Backfill

The first, complete collection of everything since the cutoff date.

Incremental run

A later run that only fetches what's new or changed. Safe to run as often as you like.

SHA-256

A fingerprint of a file's exact contents. Two files with the same SHA-256 are byte-for-byte identical.

Revision

A new version of a file that changed on the site. Saved as .rev2, .rev3, and so on.

Zip member

A file that was inside a downloaded .zip. It gets its own row, linked by parent_zip.

robots.txt

A site's published rules for automated visitors. The collector reads and obeys PJM's.

Throttle

The deliberate pause between requests that keeps the crawl polite.

12. Getting help

Questions, a missing file, or an idea? Use the **Feedback** tab on any page, or message Poornima Ramakrishnan on Upwork. Replies usually come within a few hours during US Eastern business hours, often sooner.