

# archivers.space

Information gathered during Data Rescue Events through the archivers.space app. See <http://www.ppehlab.org/datarescueworkflow#retiredworkflow> for more information.

Priority \*

Number should be 1-10

5

URL \*

<http://www.eia.gov/coal/annual/>

Agency \*

United States Department of Energy

Event

Title

Annual Coal Report

Crawled by the Internet Archive? \*

☒ Yes

☐ No

Internet Archive URL

[https://web.archive.org/web/\\*/http://www.eia.gov/coal/annual/](https://web.archive.org/web/*/http://www.eia.gov/coal/annual/)

Description

Purpose or significance of data

- Coal Production and number of mines by state, number of employees by mine, etc.
- High significance

- ☐ Do not harvest. All data is small, unstructured, and on a page crawlable by the Internet Archive.
- ☐ Page contains dynamic content (e.g., links loaded by JavaScript).
- ☐ Page contains interactive visualizations.
- ☒ Data is accessible in structured file(s) that can be directly downloaded.
- ☐ Data is accessible over FTP.
- ☐ Data is accessible using a documented public API.
- ☐ Data is only accessible using search queries in a web form.

Recommended approach to harvesting data

File formats

PDF, CSV

Estimated size in MB

Related URLs

eia.gov/coal

Where you able to capture all of the data at this URL?

☒ Yes

☐ No

Harvest method used

python scraper

Notes from Harvest

Tables in excel and pdfs all downloaded

User certified that to the best of their knowledge this is a well-checked bag that will survive out of context of the site it was harvested from.

☒ Yes

Notes from Bagging

Notes from Describe

This form was created outside of your domain.

Google Forms