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

0

URL *

https://www.energy.gov/data/downloads/open-data-catalogue

Agency *

DOE

Event

Title

Open Data Catalogue

Crawled by the Internet Archive? *

○ Yes

○ No

Internet Archive URL

https://web.archive.org/web/*/https://www.energy.gov/data/downloads/open-data-catalogue

Description

Purpose or significance of data

The files contain the Public Data Listing for the Department of Energy in .JSON format. The Public Data Listing is a machine-readable list of all the publically available datasets maintained by the Department and its program and staff offices.

☐ 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

xlsx, json

Estimated size in MB

Related URLs

www.energy.gov/data/downloads/open-data-catalogue

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

◉ Yes

○ No

Harvest method used

Direct download

Notes from Harvest

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

Bagged with Data Conservancy Packaging tool; validated before S3 upload and after S3 download.

Notes from Describe