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

6

URL *

http://apps1.eere.energy.gov/states/

Agency *

Energy Department

Event

Title

EERE: Clean Energy in My State Home Page

Crawled by the Internet Archive? *

- (●) Yes
- ( ) No

Internet Archive URL

https://web.archive.org/web/*/http://apps1.eere.energy.gov/states/

Description

Purpose or significance of data

Historical statistics on energy supply and consumption for different states, from 1960 to present. It is displayed as a whole, and broken down into residential, commercial, industrial, and transportation sectors. It also breaks the data down by renewable energy sources, including: Hydroelectric, Biomass, Geothermal, Solar, Wind power.

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

Recommended approach to harvesting data

The underlying data is available directly from the State Energy Data System (SEDS) database at http://www.eia.gov/state/seds/ as plain CSV files, per state, as well as a JSON API with an access key.

The charts are fairly straightforward highcharts.js charts, loaded into the page within iframes. However, they load data from a JSON API that would not be captured by a crawler. The most complex chart is a comparison between states, and loads the same data on every state-specific page from a JSON API (using the JSONP technique). The charts could be saved as is, and the `EIA_grapher.js` file modified to load data from an archived copy of the JSON instead of the live API. The JavaScript that generates the charts is marked as public domain and can be freely modified.

As a test I saved a chart frame for one state to my computer using the browser's Save Page As feature, as well as the response for the JSONP data request, and modified the script to load the data from a local file. The local copy of the chart was fully functional.

File formats

CSV, JSON

Estimated size in MB

1MB

Related URLs

http://apps1.eere.energy.gov/states/

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

- (●) Yes
- ( ) No

Harvest method used

Manual download of individual webpages and SEDS CSV file.

Notes from Harvest

The core of the harvest is a CSV file containing data from the State Energy Data System (SEDS) from 1960-2014. A folder of html files is also included. These files serve as templates for state-level webpages that are dynamically generated, but contain essentially the same content (where state names and data series are changed as appropriate).

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

Validated before S3 upload and after S3 download

Notes from Describe