1. Priority *

5

2. URL *

https://epa.gov/ghgreporting

3. Agency *

N/A

4. Event

5. Title

Greenhouse Gas Reporting Program (GHGRP) - US Environmental Protection Agency (EPA)

6. Crawled by the Internet Archive *

◉ Yes

7. Internet Archive URL

https://web.archive.org/web/*/https://www.epa.gov/ghgreporting/

8. Description

9. Purpose or significance of data

Includes data sets from the Greenhouse Gas Reporting Program of the US Environmental Protection Agency. Includes 2015 data highlights, facility level information, envirofacts, industrial profile, and miscellaneous sets of data. 575D2AB9-104D-491D-A8AF-7BDC94253B35

10.

☐ 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.

11. Recommended approach to harvesting data

There's a section on the page for GHGRP Data. 2015 Data highlights links to a page that has some data and graphs in boxes. More data is linked on the right hand side. At the bottom of the box, GHGRP Data Sets links back to other datasets including 2015 Data Summary Spreadsheets. Facility Level Information on GHGs Tools (FLIGHT) is an interactive GIS mapping system. Envirofacts includes searches for data in a web form. Industrial Profiles includes PDFs of data sets.

12. File formats

PDF, ZIP, XLSX

13. Estimated size in MB

14. Related URLs

575D2AB9-104D-491D-A8AF-7BDC94253B35

15. Were you able to capture all of the data at this URL?

◉ Yes

◯ No

16. Harvest method used

wget script with a file containing names of downloaded files.

17. Notes from Harvest

Downloaded recent (2010-2014) Greenhouse Gas Reporting Program (GHGRP) data, which collects Greenhouse Gas (GHG) data from large emitting facilities, suppliers of fossil fuels and industrial gases that result in GHG emissions when used, and facilities that inject carbon dioxide underground. Could not harvest the Envirofacts datasets (links to another site altogether), Added Envirofacts URL to researching queue. Grabbed everything else.

18. 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

19. Notes from Bagging

Bagged with Data Conservancy Packaging Tool; Validated before S3 upload and after S3 download.

20. Notes from Describe

Used description specifically from https://www.epa.gov/ghgreporting/ghg-reporting-program-data-sets