

Responder 29



19 - Anonymous



04:39

Time to complete



1. Priority *

1

2. URL *

<https://epa.gov/toxics-release-inventory-tri-program/environmental-toxicity-data-tri-listed-chemicals>

3. Agency *

EPA

4. Event

5. Title

Toxics Release Inventory (TRI) Program - TRI-Listed Chemicals

6. Crawled by the Internet Archive *

☒ Yes

7. Internet Archive URL

https://web.archive.org/web/*/https://epa.gov/toxics-release-inventory-tri-program/environmental-toxicity-data-tri-listed-chemicals

8. Description

9. Purpose or significance of data

TRI List of Reportable Chemicals spreadsheets for 2015-2017 data (.pdf of 1997-2016)

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

12. File formats

13. Estimated size in MB

14. Related URLs

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

☒ Yes

☐ No

16. Harvest method used

17. Notes from Harvest

XLSX files from three pages: https://www.epa.gov/sites/production/files/2016-12/tri_chemical_list_for_ry15_11_5_2015_0_2_0.xlsx
https://www.epa.gov/sites/production/files/2016-12/tri_chemical_list_for_ry16_12_8_2016_0_0_0.xlsx
https://www.epa.gov/sites/production/files/2016-12/tri_chemical_list_for_ry17_12_8_2016_0_0_0.xlsx

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