Responder 29

2 - Anonymous

04:24

Time to complete

1. Priority *

1

2. URL *

http://epa.gov/dwssa

3. Agency *

N/A

4. Event

5. Title

Sole Source Aquifers for Drinking Water

6. Crawled by the Internet Archive *

- (●) Yes

7. Internet Archive URL

https://web.archive.org/web/*/http://epa.gov/dwssa

8. Description

9. Purpose or significance of data

An overview of EPA-defined sole source aquifers for drinking water

10.

- [ ] 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.
- [ ] 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 link to an ArcGIS map here
https://epa.maps.arcgis.com/apps/webappviewer/index.html?
id=9ebb047ba3ec41ada1877155fe31356b

12. File formats

13. Estimated size in MB

14. Related URLs

BAE2F5AF-369B-4894-8236-245680752518

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

- (●) Yes
- ( ) No

16. Harvest method used

esri2geojson

17. Notes from Harvest

In the visualizer app, if you click the "Layer List" button at top right, you can click the three dots icon on each entry to get a details menu. You can then click "Show Item Details" or "Description". For the former, you'll be taken directly to a "ArcGIS REST Services Directory". For the former, you'll get a pretty overview page--find the "Layers" section for a link to the REST page. On the REST page, if the URL ends in FeatureServer, you can use esri2geojson to download the data-- append /0, /1, etc to the url to get REST endpoints for the first, second, etc. layers of the feature server. If the URL ends in MapServer, it seems that the only export option available is as a KMZ file which is simply a wrapper around link to a web-hosted resource such as https://geodata.epa.gov/arcgis/rest/services/OEI/EPA_Locations/MapServer/export . Regardless, I download these as well. I believe that the main data is in what comes from the FeatureServers. This is scripted in dl.py and arcgisGet.py

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 (v.1.0.5); Validated before S3 upload and after S3 download.

20. Notes from Describe