

Responder 29



9 - Anonymous



02:46



Time to complete

1. Priority \*

10

2. URL \*

<http://epa.gov/region9/air/maps>

3. Agency \*

N/A

4. Event

5. Title

EPA Trends/Pacific Southwest, Region 9: Air Quality Analysis

6. Crawled by the Internet Archive \*

☒ Yes

## 7. Internet Archive URL

[https://web.archive.org/web/\\*/http://epa.gov/region9/air/maps](https://web.archive.org/web/*/http://epa.gov/region9/air/maps)

## 8. Description

## 9. Purpose or significance of data

Documents data of the air quality and geographic maps of the states located in the US Pacific Southwest.

## 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

Most useful data available in pdf, zip and kml files linked from web pages but not appearing in WayBack machine (perhaps due to frames). Manually retrieve zip files and if possible, link them to the wayback machine corresponding pages?

## 12. File formats

ZIP, KML, PDF

## 13. Estimated size in MB

200

## 14. Related URLs

<http://epa.gov/region9/air/maps>

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

☒ Yes

☐ No

## 16. Harvest method used

```
wget -r -k -K -np -p https://www3.epa.gov/region9/air/maps/
```

17. Notes from Harvest

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