

Responder 29



11 - Anonymous



02:38



Time to complete

1. Priority \*

10

2. URL \*

[http://epa.gov/region09/air/maps/r9\\_o38hr1997.html](http://epa.gov/region09/air/maps/r9_o38hr1997.html)

3. Agency \*

N/A

4. Event

5. Title

California Ozone (O3) Attainment Designations (1997 8-hour standard)

6. Crawled by the Internet Archive \*

☒ Yes

## 7. Internet Archive URL

[https://web.archive.org/web/\\*/http://epa.gov/region09/air/maps/r9\\_o38hr1997.html](https://web.archive.org/web/*/http://epa.gov/region09/air/maps/r9_o38hr1997.html)

## 8. Description

## 9. Purpose or significance of data

Geographic data showing regions in California by their ozone attainment designations in 1997. The data are stored in two formats: as a zip file that stores ArcGIS-readable data (should have been crawlable) and as a kmz file that may not be a format that the crawler would have saved. That file can be manually downloaded.

## 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

Directly save the file <https://www3.epa.gov/region9/air/maps/kmz/epa-reg9-97-8hr-o3-naaqs-designs.kmz>

## 12. File formats

kmz

## 13. Estimated size in MB

1.7

## 14. Related URLs

[http://epa.gov/region09/air/maps/r9\\_o38hr1997.html](http://epa.gov/region09/air/maps/r9_o38hr1997.html)

## 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/r9\\_o38hr1997.html](https://www3.epa.gov/region9/air/maps/r9_o38hr1997.html)

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