

archivers.space

Information gathered during Data Rescue Events through the archivers.space app. See <http://www.ppehlab.org/datarescueworkflow#retiredworkflow> for more information.

Priority *

Number should be 1-10

1

URL *

https://www.glerl.noaa.gov//metdata/status/status_archive/

Agency *

National Oceanic and Atmospheric Administration

Event

Title

Index of /metdata/status/status_archive

Crawled by the Internet Archive? *

Yes

No

Internet Archive URL

<https://www.archivers.space/urls/DB4540E2-81B3-418A-B5BB-C222F24F6FE5>

Description

Purpose or significance of data

hourly measurements of atmospheric and oceanic data

- 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.

Recommended approach to harvesting data

just loop through and download html files (they aren't archived)

File formats

HTML

Estimated size in MB

100MB

Related URLs

https://www.glerl.noaa.gov//metdata/status/status_archive/

Where you able to capture all of the data at this URL?

Yes

No

Harvest method used

recursive wget

Notes from Harvest

very straightforward, but response was very slowly, probably why internet archive does not yet have it.

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

Notes from Bagging

Notes from Describe

This form was created outside of your domain.

Google Forms