

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

3

URL *

<http://www.st.nmfs.noaa.gov/nauplius/>

Agency *

National Oceanic and Atmospheric Administration

Event

Title

NAUPLIUS: Advanced Ecosystem Data Products

Crawled by the Internet Archive? *

☒ Yes

☐ No

Internet Archive URL

<https://www.archivers.space/urls/95D74DF1-5B1E-4B84-BFF7-2F40FFA92721>

Description

Purpose or significance of data

Plankton biomass 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

Use wget, starting at <http://www.st.nmfs.noaa.gov/copepod/atlas/index-atlas.html> and going at least two levels deep.

File formats

TXT

Estimated size in MB

Related URLs

<http://www.st.nmfs.noaa.gov/nauplius/>

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

☒ Yes

☐ No

Harvest method used

wget --recursive --level=2 --accept html,txt

<http://www.st.nmfs.noaa.gov/copepod/atlas/index-atlas.html>

Notes from Harvest

Most of the data is in www.st.nmfs.noaa.gov/copepod/atlas/data_src/.

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