

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

2

URL *

<http://www.nssl.noaa.gov/projects/wof/documents/workshop2013/index.php>

Agency *

National Oceanic and Atmospheric Administration

Event

Title

NSSL: Warn-on-Forecast : 2013 Workshop

Crawled by the Internet Archive? *

☒ Yes

☐ No

Internet Archive URL

Description

Purpose or significance of data

Powerpoint slides from a 2013 conference on providing warnings for severe weather from the National Severe Storms Library

- ☐ 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

should be pretty easy to use the w-get loop

File formats

pdf, pptx, pdf, mov

Estimated size in MB

Related URLs

<http://www.nssl.noaa.gov/projects/wof/documents/workshop2013/index.php>

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

☒ Yes

☐ No

Harvest method used

Webserver prevents directory listing, so not possible to scrape presentations using wget. Wrote a 10 line python script to parse the HTML and pull out link urls, ran wget on the resulting urls. Python script available in tools/

Notes from Harvest

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