

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.goes-r.gov/users/2015-OCONUS.html>

Agency *

National Oceanic and Atmospheric Administration

Event

Title

Past Conference - 2015 GOES-R/JPSS OCONUS Satellite Proving Ground Meeting

Crawled by the Internet Archive? *

☒ Yes

☐ No

Internet Archive URL

<https://www.archivers.space/urls/7E7E2318-5793-4F9C-931D-BE6BCA4A6BC6>

Description

Purpose or significance of data

Presentations from a meeting related to GOES-R weather satellite

- ☐ 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

File formats

PDF, PPT

Estimated size in MB

Related URLs

<http://www.goes-r.gov/users/2015-OCONUS.html>

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

☒ Yes

☐ No

Harvest method used

Shell command

Notes from Harvest

The following shell command was used to suss out and download all .pptx URLs on the main page of the GOES-R website:

```
curl -s http://www.goes-r.gov/users/2015-OCONUS.html \  
| grep '\.pptx' \  
| sed 's/.*href="\([^"]*\)".*/\1/' \  
| sed 's/ /\ /g' \  
| uniq \  
| awk '{print "http://www.goes-r.gov" $0}' \  
| xargs wget
```

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