

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 *

http://www.goes-r.gov/users/AMS-2016_BroadcastersShortCourse.html

Agency *

National Oceanic and Atmospheric Administration

Event

Title

AMS 2016 Broadcasters Short Course

Crawled by the Internet Archive? *



Yes



No

Internet Archive URL

<https://www.archivers.space/urls/8073A855-3499-44E3-9C9E-CAD3DD8FEE64>

Description

Purpose or significance of data

Information about the tools on GOES-R (now GOES-16) weather satellites

- ☐ 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, pptx

Estimated size in MB

Related URLs

http://www.goes-r.gov/users/AMS-2016_BroadcastersShortCourse.html

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

☒ Yes

☐ No

Harvest method used

python3 script using beautifulsoup, wget and urllib

Notes from Harvest

the metadata json seed provided by the zip starter was missing the id_agency, id_subagency, id_org, id_suborg tags, as well as containing a few tags which did not match the format in <https://github.com/datarefuge/workflow/blob/master/harvesting.md>, nor in <https://github.com/edgi-govdata-archiving/workflow/blob/master/id-json.md>

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.

