

# 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/2014-GLM.html>

---

## Agency \*

National Oceanic and Atmospheric Administration

---

## Event

---

## Title

2014 GLM Science Meeting Presentations

---

## Crawled by the Internet Archive? \*

☒ Yes

☐ No

## Internet Archive URL

<https://www.archivers.space/urls/AFFDBD94-054C-477E-82CE-C5D59BB57EB8>

---

## Description

---

### Purpose or significance of data

These are presentations for the GOES mission. GOES are geostationary satellites with a bunch of instrumentation for climate and weather observations. While the presentations themselves are of lower significance, the data when it comes online will be important for weather forecasting and furthering understanding of climate science, space science, etc. It looks like the GOES-R mission has launched and the operational data are not online yet. This particular page has .pdf and .pptx files that document meetings, etc. There are other pages that have image files that show preliminary, non-operational data. The operational data would be important to archive once they are online.

---

- ☐ 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

On this particular page, the files are small and can be directly downloaded. (It is not structured data). It doesn't look like any of the scientific data are here. The mission, after launched, is now called GOES-16. Can't find a place where those data are available.

---

## File formats

pdf, jpg, pptx

---

## Estimated size in MB

---

## Related URLs

<http://www.goes-r.gov/users/2014-GLM.html>

---

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

☒ Yes

☐ No

## Harvest method used

Downloaded all PDF files with a shell command

---

## Notes from Harvest

Used the following shell command to suss out and download PDFs linked to from the <http://www.goes-r.gov/users/2014-GLM.html> website.

```
curl -s http://www.goes-r.gov/users/2014-GLM.html \  
| grep '\.pdf' \  
| sed 's/.*href="\([^"]*\)".*/\1/' \  
| uniq \  
| awk '{print "http://www.goes-r.gov" $1}'
```

---

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