The STAR Central Data Repository (SCDR): A Scientist's Guide

Bob Kuligowski, STAR / SMCD / EMB

Chair, STAR ITAC (IT Advisory Committee) and DMG (Data Management Group)

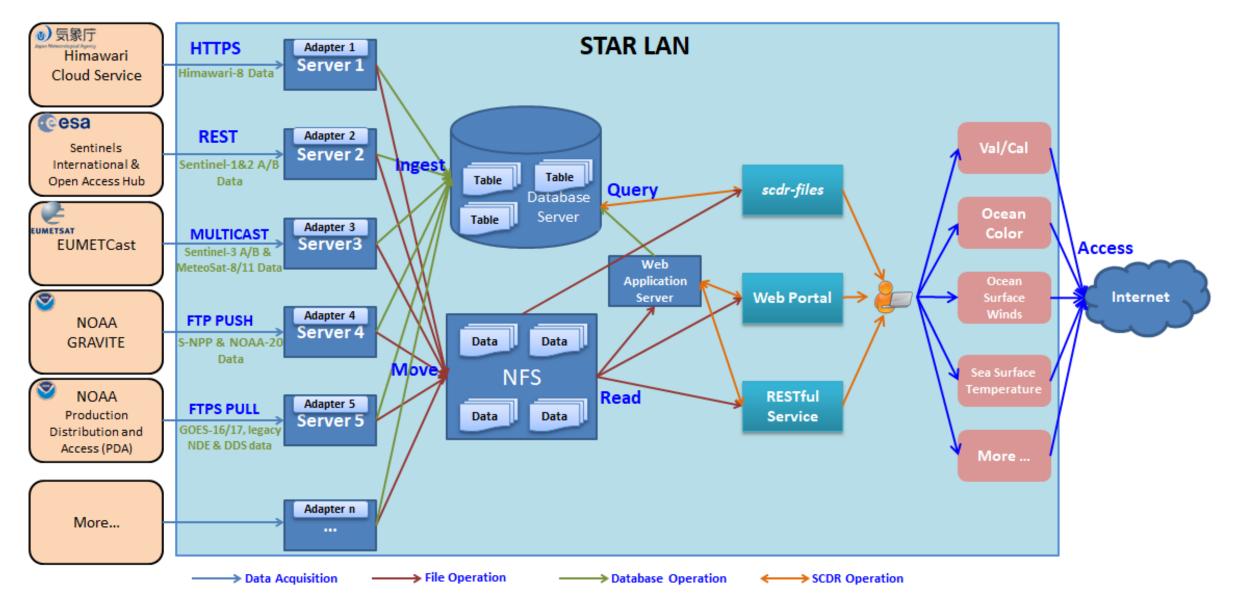
Outline

- What Is SCDR and Why Do We Have It?
- How Do I Use SCDR?
- How Are New Data Sets Added to SCDR?

What is SCDR? SCDR Facts

- SCDR = STAR Central Data Repository
- A 1.5-Petabyte central repository of satellite and selected ground and model data, including:
 - Polar operational satellites (POES, JPSS, DMSP)
 - Geostationary operational satellites (GOES, Meteosat, Himawari)
 - Non-operational, non-NOAA satellites (Terra/Aqua, GPM, Sentinel)
 - GFS and RAP model forecasts
- Because of space limitations, generally contains a 1-year repository of L1b data and a 90 to 120-day repository of L2 data
- SCDR is accessible only from the STAR LAN!
- SCDR is managed by the STAR Data Management Group (DMG)

What is SCDR? SCDR Architecture



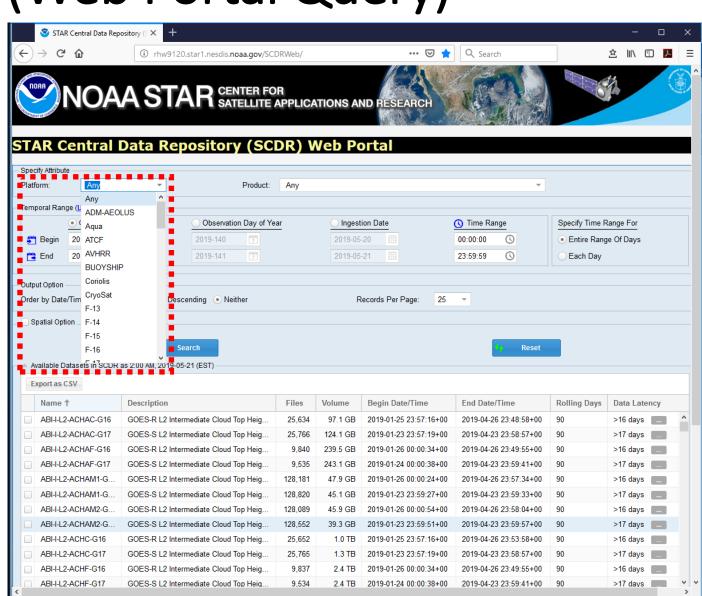
Why Do We Have SCDR?

- Avoids wasting disk space with multiple repositories of the same dataset in multiple locations
- Avoids redundant data pulls by serving as a single point of entry for datasets into STAR from e.g., PDA and GRAVITE

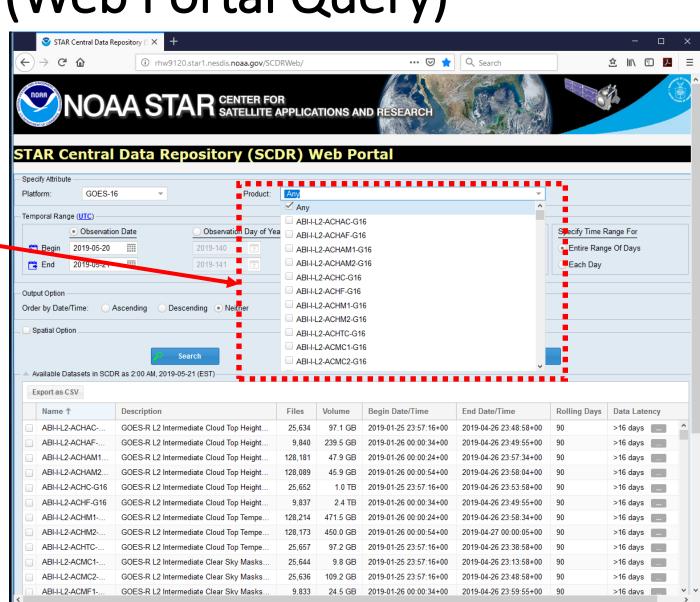
How Do I Use SCDR?

- The SCDR database can be queried from a Web portal, a Linux command, or an API.
- Web portal: http://rhw9120.star1.nesdis.noaa.gov/SCDRWeb/

Select the satellite platform or model you want

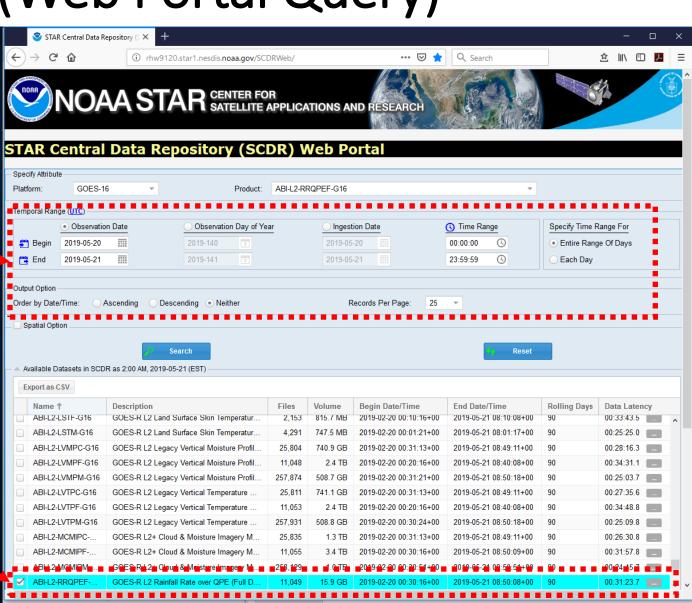


Select the specific data set you want

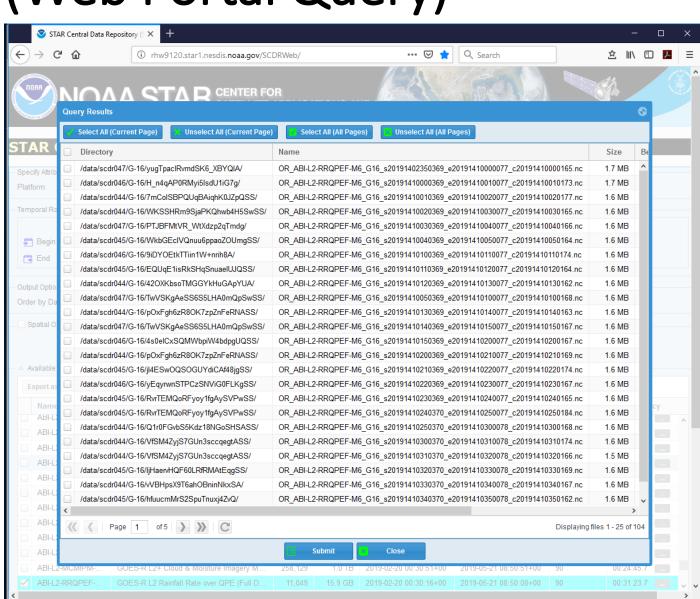


Select the range of dates and times you want and your output options

SCDR will give the time range of the available data and the total volume

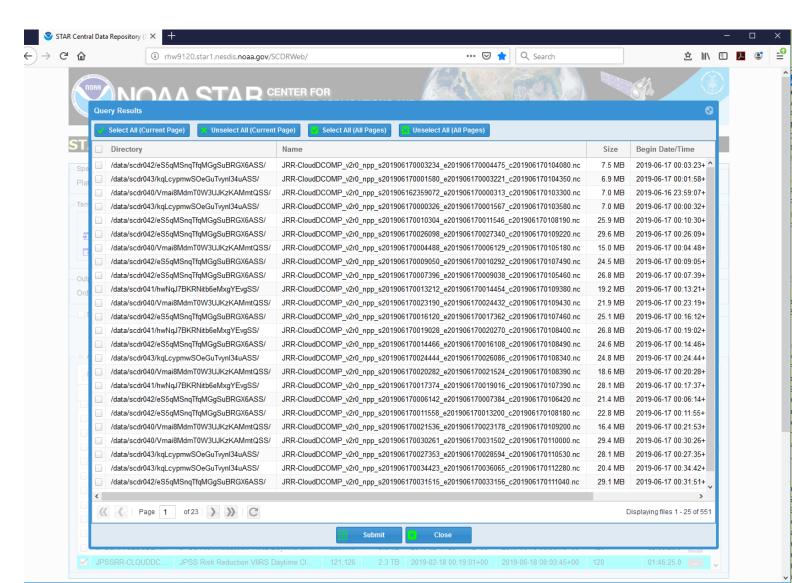


Query returns a list of files, pathnames, sizes, and start/end times (need to scroll right to see the rest)

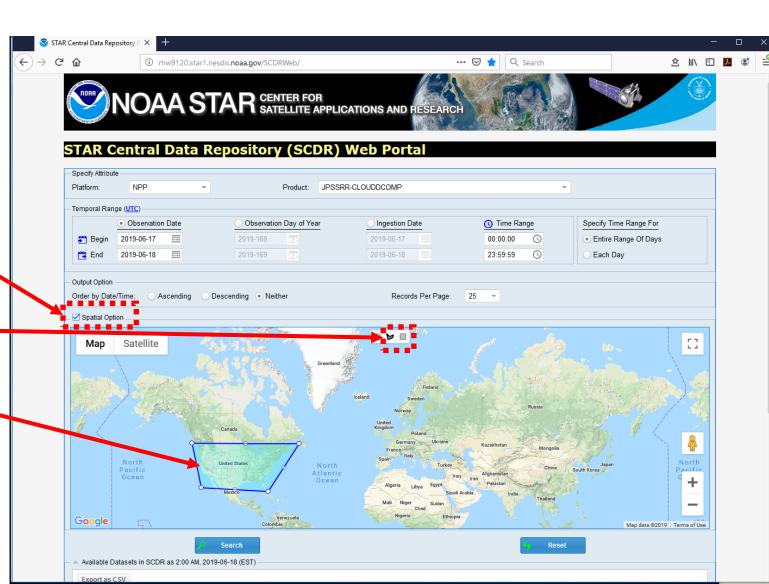


For selected datasets, you can also search for data in a particular location by clicking the "spatial option" check box and using the tools to draw a search polygon.

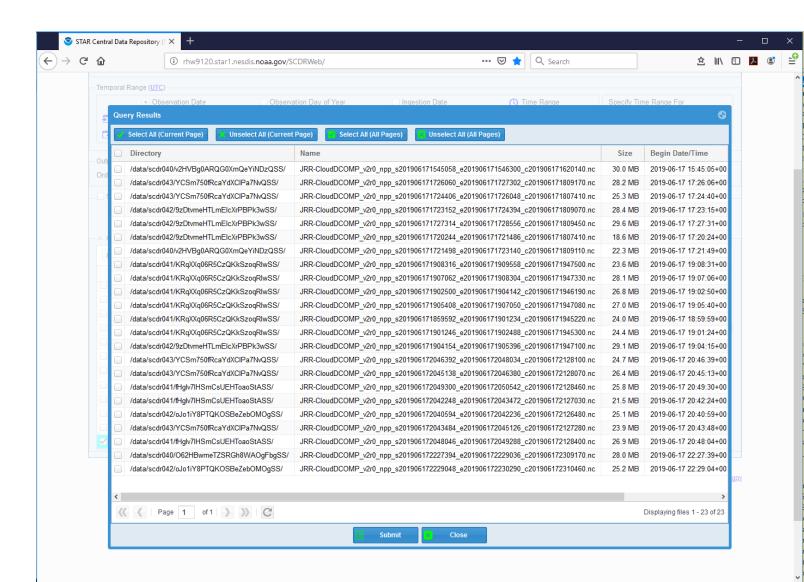
For example, a search for NPP DCOMP from for all of 17 June 2019 returns 551 entries that are all over the globe:



Selecting the "Spatial Option" check box and drawing a polygon over the CONUS using the polygon tool:



Now only 23 granules over the CONUS are returned!



- The Linux command /data/starfs1/bin/scdr-files can be used for database queries from the Linux command line or a script
- /data/starfs1/bin/scdr-files -h(elp) returns documentation.
- Some useful options (but not all of them):
 - -available ← summary report of EVERYTHING in the database
 - -t <type> \leftarrow collection type (from the 1st column in the summary report)
 - -stime "YYYY-MM-DD at HH:MM:SS UTC" ← start time of query
 - -etime "YYYY-MM-DD at HH:MM:SS UTC" ← end time of query
 - -satname "satellite name" ← satellite name
 - -a "POLYGON ((LAT.1 LON.1, LAT.2 LON.2, LAT.3 LON.3, ...LAT.1, LON.1))" ← granules only within specified polygon

GOES-16 ABI Level 2 RRQPEF product
Search period starts at 1500 UTC 21 May 2019

/data/starfs1/bin/scdr-files -t ABI-L2-RRQPEF-G16

-stime "2019-05-21 15:00 UTC" -etime "2019-05-21

15:59 UTC"

Search period ends at 1559 UTC 21 May 2019

```
bob.kuligowski@rhw1061:/data/data081/bobk/ABI/Scripts — ABI-L2-RRQPEF-G16 -stime "2019-05-21 15:00 UTC" -etime "2019-05-21 15:59 UTC" /data/scdr044/G-16/uu8ifNlsSFa5E7IvUZNlMwSS/OR_ABI-L2-RRQPEF-M6_G16_s20191411500374_e20191411500082_c20191411500162.nc /data/scdr046/G-16/4EGD5RIDQG60PpojaWZYLwSS/OR_ABI-L2-RRQPEF-M6_G16_s20191411500374_e20191411510082_c20191411510164.nc /data/scdr046/G-16/4EGD5RIDQG60PpojaWZYLwSS/OR_ABI-L2-RRQPEF-M6_G16_s20191411510374_e20191411520082_c20191411520215.nc /data/scdr044/G-16/67ASMNetRLusqMzpS5W66ASS/OR_ABI-L2-RRQPEF-M6_G16_s20191411520374_e20191411530082_c20191411530165.nc /data/scdr045/G-16/sSe6TyrdSj6CL2W3SvAbNASS/OR_ABI-L2-RRQPEF-M6_G16_s20191411530374_e20191411540082_c20191411540167.nc /data/scdr047/G-16/Sk8BS4HfQ36PguIBc8G05wSS/OR_ABI-L2-RRQPEF-M6_G16_s20191411540374_e20191411550082_c20191411550166.nc /data/scdr045/G-16/sSe6TyrdSj6CL2W3SvAbNASS/OR_ABI-L2-RRQPEF-M6_G16_s20191411550374_e20191411600082_c20191411600164.nc bob.kuligowski@rhw1061:
```

- You can then copy the files you need to your directory (or softlink them if you're not going to use them more than once)
- To automate this, have a script make a query, dump the results to a text file, then use the text file as a file list for copying.

Data from NPP only

Search period starts at O000 UTC 17 June 2019

/data/starfs1/bin/scdr-files -t JPSSRR-CLOUDDCOMP
-satname "npp" -stime "2019-06-17 00:00 UTC"

-etime "2019-06-17 23:59 UTC" -a "POLYGON ((25.0 - 125.0, 49.0 - 125.0, 49.0 - 65.0, 25.0 - 65.0, 25.0 - 65.0)"

Search period starts at 2359 UTC 17 June 2019

Search for granules intersecting 25-50°N, 125-65°W.

- SCDR RESTful Web service allows wget or curl or similar commandline tools to make database queries in place of the scdr-files command
- Documentation is at http://rhw9120.star1.nesdis.noaa.gov/SCDRWeb/help.html
- It allows output in HTML, XML or JSON format in addition to plain text:
 - Content-Type: text/plain (default) ← plain text
 - Content-Type: text/html ← HTML
 - Content-Type: application/xml ← XML
 - Content-Type: application/json ← JSON

- The option syntax for using the API is a bit different from scdr-files:
 - type=<type> ←dataset name in SCDR database
 - stime=YYYY-MM-DD**T**HH:MM:SS ← start time of query
 - etime=YYYY-MM-DD**T**HH:MM:SS ← end time of query
 - satname=<satname> ←satellite name in SCDR database
 - since=YYYY-M(M)-DD**T**HH:MM:SS ← query from start time until now
 - area=POLYGON (LAT.1 LON.1, LAT.2 LON.2 LAT.3 LON.3,...LAT.1 LON.1) ← polygon encompassing search area (any granules that even partially intersect the polygon will be selected)

Sample query of file types in SCDR using wget:

Write output to result.txt (default is to write to a file called list)

wget --header "Content-Type: text/plain" -O ./result.txt

"http://rhw9120.star1.nesdis.noaa.gov/SCDRWeb/service/scdr-files/list"

url containing the list of SCDR file types

bob.kuligowski@rhw1061:/data/data081/bobk/ABI/Scripts

Sample query of file types in SCDR using wget:

```
bob.kuligowski@rhw1061: wget --header "Content-Type: text/plain" -O ./result.txt "http://rhw9120.starl.nesdis.noaa.gov/SCDRWeb/service/scdr-files/list"
--2019-06-11 18:40:54-- http://rhw9120.starl.nesdis.noaa.gov/SCDRWeb/service/scdr-files/list
Resolving rhw9120.starl.nesdis.noaa.gov... 10.2.7.81
Connecting to rhw9120.starl.nesdis.noaa.gov|10.2.7.81|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/plain]
Saving to: "./result.txt"
                                                                                                                                      ] 119,536
   [ <=>
2019-06-11 18:40:54 (97.7 MB/s) - "./result.txt" saved [119536]
bob.kuligowski@rhwl061: cat result.txt
Type, Description, Files, Volumn (byte), Begin Date, End Date, Rolling Days
ABI-I-L2-ACHAC-G16, "GOES-R L2 Intermediate Cloud Top Height (CONUS) product", 25851, 105508420812, "2019-02-10 23:57:14+00", "2019-05-11 23:59:06+00", 90
ABI-I-L2-ACHAC-G17, "GOES-S L2 Intermediate Cloud Top Height (CONUS) product", 25795, 130210260929, "2019-02-10 23:57:19+00", "2019-05-11 23:58:57+00", 90
ABI-I-L2-ACHAF-G16, "GOES-R L2 Intermediate Cloud Top Height full disk product", 10650, 281226852996, "2019-02-11 00:00:31+00", "2019-05-12 00:00:03+00", 90
ABI-I-L2-ACHAF-G17, "GOES-S L2 Intermediate Cloud Top Height full disk product", 10154, 278500792928, "2019-02-11 00:00:38+00", "2019-05-11 23:59:41+00", 90
ABI-I-L2-ACHAM1-G16, "GOES-R L2 Intermediate Cloud Top Height mesoscale mode 1 product", 129170, 52729327045, "2019-02-11 00:00:22+00", "2019-05-11 23:59:42+00", 90
ABI-I-L2-ACHAM1-G17, "GOES-S L2 Intermediate Cloud Top Height mesoscale mode 1 product", 128965, 46847868788, "2019-02-10 23:59:27+00", "2019-05-11 23:59:33+00", 90
ABI-I-L2-ACHAM2-G16, "GOES-R L2 Intermediate Cloud Top Height mesoscale mode 2 product", 129123, 50532576827, "2019-02-10 23:59:52+00", "2019-05-11 23:59:12+00", 90
ABI-I-L2-ACHAM2-G17," GOES-S L2 Intermediate Cloud Top Height mesoscale mode 2 product ",128700,41820182798,"2019-02-10 23:59:51+00","2019-05-11 23:59:57+00",90
ABI-I-L2-ACHC-G16, "GOES-R L2 Intermediate Cloud Top Height (CONUS) product", 25851, 1079509849064, "2019-02-10 23:57:14+00", "2019-05-11 23:59:06+00", 90
ABI-I-L2-ACHC-G17, "GOES-S L2 Intermediate Cloud Top Height (CONUS) product", 25795, 1369679423644, "2019-02-10 23:57:19+00", "2019-05-11 23:58:57+00", 90
ABI-I-L2-ACHF-G16, "GOES-R L2 Intermediate Cloud Top Height full disk product", 10651, 2878311056096, "2019-02-11 00:00:31+00", "2019-05-12 00:00:03+00", 90
ABI-I-L2-ACHF-G17, "GOES-S L2 Intermediate Cloud Top Height full disk product", 10152, 2833786435120, "2019-02-11 00:00:38+00", "2019-05-11 23:59:41+00", 90
ABI-I-L2-ACHM1-G16, "GOES-R L2 Intermediate Cloud Top Temperature mesoscale mode 1 product", 129171, 519609815439, "2019-02-11 00:00:22+00", "2019-05-11 23:59:42+00", 90
ABI-I-L2-ACHM1-G17, "GOES-S L2 Intermediate Cloud Top Temperature mesoscale mode 1 product", 128963, 443468169619, "2019-02-10 23:59:27+00", "2019-05-11 23:59:33+00", 90
ABI-I-L2-ACHM2-G16, "GOES-R L2 Intermediate Cloud Top Temperature mesoscale mode 2 product", 129123, 495698613595, "2019-02-10 23:59:52+00", "2019-05-11 23:59:12+00", 90
```

• Same GOES-16 RRQPE example as before, but using wget:

```
Write output as
url for querying
                                          Write output to SCDRlist.txt
                    plain (ASCII) text
specific datasets
                                       (default is to write to a file called list)
       -header "Content-Type: text/plain" -0
                                                     ./SCDRlist.txt
"http://rhw9120.star1.nesdis.noaa.gov/SCDRWeb/service/scdr-
files/query?type=ABI-L2-RRQPEF-G16&stime=2019-5-21T15:00:0
&etime=2019-5-21T15:59:\00"
                                              Start query at 1500
                     Get G16 RRQPE files
 End query at 1559
                                               UTC 21 May 2019
 UTC 21 May 2019
```

• Same GOES-16 RRQPE example as before, but using wget:

(Who wants to type all that, anyway? So...)

• Same GOES-16 RRQPE example again, but using wget in a script:

```
bob.kuligowski@rhw1061:/data/data081/bobk/ABI/Scripts
bob.kuligowski@rhwl061: cat test wget.sh
url=http://rhw9120.starl.nesdis.noaa.gov/SCDRWeb/service/scdr-files/guery
type=ABI-L2-RRQPEF-G16
stime=2019-5-21T15:00:00
etime=2019-5-21T15:59:00
#get --header "Content-Type: text/plain" -O ./SCDRlist.txt "${url}?type=${type}&stime=${stime}&etime=${etime}"
cat SCDRlist.txt
bob.kuligowski@rhw1061:
bob.kuligowski@rhwl061: ./test wget.sh
--2019-06-11 18:56:53-- http://rhw9120.starl.nesdis.noaa.gov/SCDRWeb/service/scdr-files/query?type=ABI-L2-RRQPEF-G16&sti
me=2019-5-21T15:00:00&etime=2019-5-21T15:59:00
Resolving rhw9120.starl.nesdis.noaa.gov... 10.2.7.81
Connecting to rhw9120.starl.nesdis.noaa.gov|10.2.7.81|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 833 [text/plain]
Saving to: "./SCDRlist.txt"
2019-06-11 18:56:53 (68.7 MB/s) - "./SCDRlist.txt" saved [833/833]
/data/scdr044/G-16/uu8ifNlsSFa5E7IvUZN1MwSS/OR_ABI-L2-RRQPEF-M6_G16_s20191411450374_e20191411500082_c20191411500162.nc
/data/scdr046/G-16/4EGD5RIDQG60PpojaWZYLwSS/OR ABI-L2-RRQPEF-M6 G16 s20191411500374 e20191411510082 c20191411510164.nc
/data/scdr046/G-16/4EGD5RIDQG60PpojaWZYLwSS/OR ABI-L2-RRQPEF-M6 G16 s20191411510374_e20191411520082_c20191411520215.nc
/data/scdr044/G-16/67ASMNetRLusqMzpS5W66ASS/OR ABI-L2-RRQPEF-M6 G16 s20191411520374 e20191411530082 c20191411530165.nc
/data/scdr045/G-16/sSe6TyrdSj6CL2W3SvAbNASS/OR ABI-L2-RRQPEF-M6 G16 s20191411530374_e20191411540082_c20191411540167.nc
/data/scdr047/G-16/Sk8BS4HfQ36PguIBc8G05wSS/OR ABI-L2-RRQPEF-M6 G16 s20191411540374 e20191411550082 c20191411550166.nc
/data/scdr045/G-16/sSe6TyrdSj6CL2W3SvAbNASS/OR ABI-L2-RRQPEF-M6 G16 s20191411550374 e20191411600082 c20191411600164.nc
bob.kuligowski@rhw1061:
```

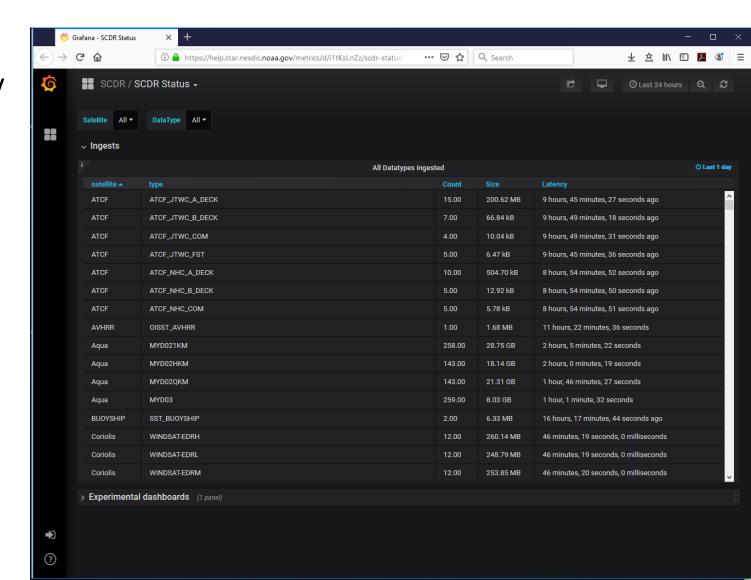
So Which Approach Should I Use?

- scdr-files currently has more options than the RESTful API
- The longer-term plan is to focus development on the RESTful API and transition scdr-files to alias to it
 - This avoids parallel development paths (and the resulting redundant work)
 - It also minimizes the need to re-work existing scripts that use scdr-files
- Recommendations:
 - Try the RESTful API, but if it doesn't meet your needs, feel free to use (or keep on using) scdr-files
 - Give feedback on what features you would like to see added!

How Do I Find Out the Latency of SCDR Data?

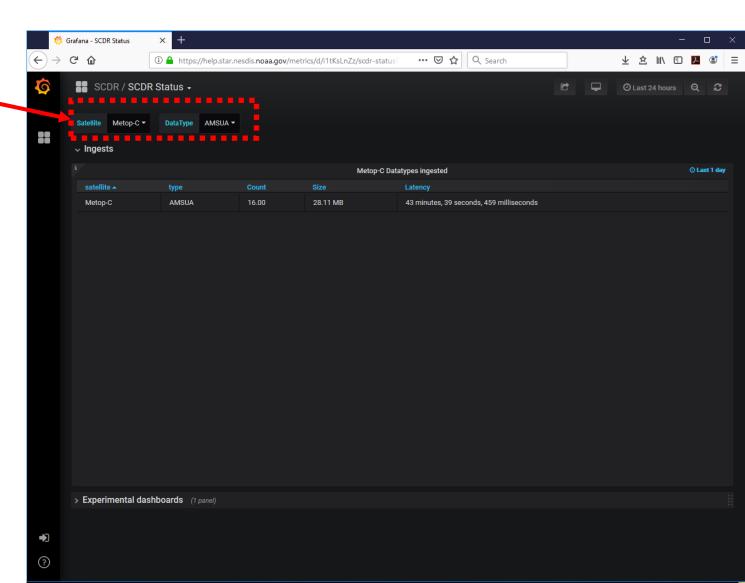
 You can use any of the methods on the previous sides to indirectly find out, but this dashboard provides a much more direct method:

https://help.star.nesdis.noaa.gov
/metrics/d/i1tKsLnZz/scdrstatus?orgId=1



How Do I Find Out the Latency of SCDR Data?

 You can also select specific satellites / data types to narrow down the list:



How Do I Use SCDR Politely?

- Really large queries significantly degrade database performance
- The scdr-files command has some built-in limits:
 - No requests without time bounds
 - Maximum time interval of 30 days
 - Requires specific data types (or file names, if known)
- Be as specific as possible in your requests, especially automated ones!

How Are New Data Sets Added to SCDR?

- The STAR Data Management Group (chaired by yours truly) manages
 SCDR ingest and storage
- To add a new data set to SCDR, fill out a SCDR <u>Data Storage Request</u> <u>Form</u> (Google Survey) at least 24 hours before the DMG meeting at which it will be voted on (2nd and 4th Wednesday of every month).
 - Information requested includes justification for the request, expected repository duration and size, and data sources.

Final thoughts...

- For any SCDR-related issues or questions, submit a STAR IT Help Desk Ticket or contact Weiguo Han at Weiguo.Han@noaa.gov
- We're always looking for ways to improve SCDR! I'll be sending out a survey soon asking for your thoughts on this.

Questions?

