

STAR Central Data Repository (SCDR): An Integrated and Effective Framework for Satellite Data Acquisition and Dissemination

Weiguo Han¹, Joseph Brust²

¹UCAR VSP at NOAA/NESDIS/STAR, ²NOAA/NESDIS/STAR
5830 University Research Court, College Park, MD 20740

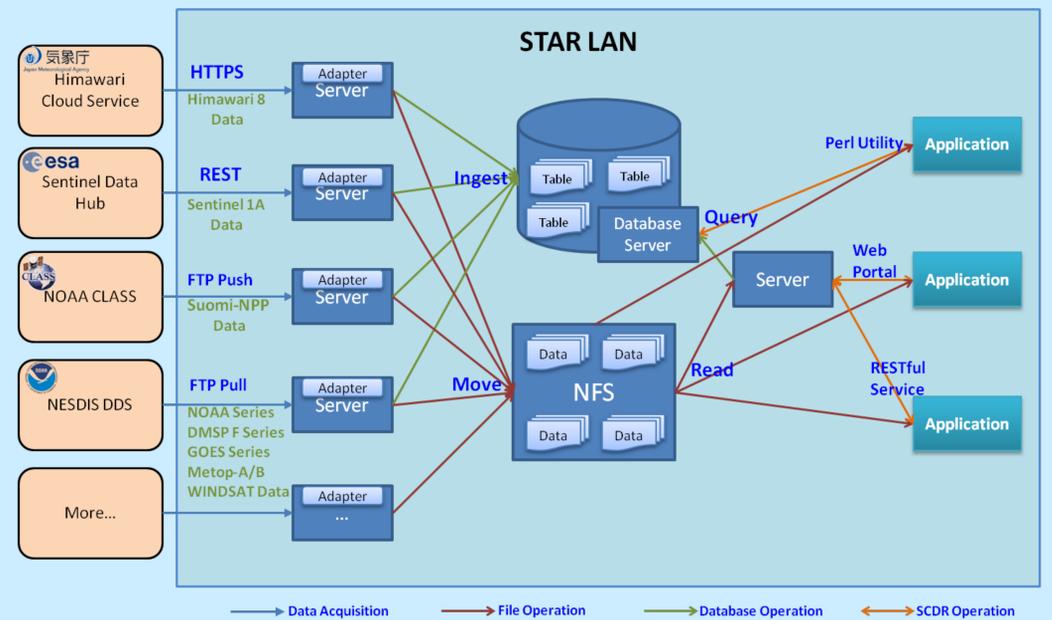
Abstract

To support near and long-term goals of the Center for Satellite Applications and Research (STAR) of the National Oceanic and Atmospheric Administration (NOAA), an integrated and effective data repository framework titled STAR Central Data Repository (SCDR) is built to provide a stable, reliable, and continually available data source, with various kinds of near real-time satellite datasets, to research and product development teams for calibration, validation, simulation, production, and monitoring activities.

In the SCDR, custom programs and utilities are developed to fetch, manage, and locate the satellite data files. The large amount of satellite and ancillary data are obtained from various providers through high speed internet connections. The management program distributes these files evenly to disks within the Network File System (NFS) environment that can be directly accessed by STAR users internally. This program retrieves metadata information (like satellite, data type, size, begin and end date/time, etc) from the files, with different formats, and ingests them into the partitioned tables in a core database. It is also responsible for managing disk space and removing the outdated files. Moreover, SCDR offers multiple easy and consistent interfaces to obtain satellite data of interest. A utility named *scdr-files* with multiple options is implemented to locate files of interest stored in the SCDR. In addition, a Web portal is built for internal users to search data of interest interactively and intuitively. And a lightweight RESTful Web service is developed as another option to list available datasets and search the specific files. This service can be easily utilized by command-line tools (such as *wget* or *curl*) or other programs across operating systems. SCDR addresses satellite data requirements and data management needs of STAR researchers and scientists. It greatly saves their time on data collecting, storing, and searching, and significantly reduces data duplication, data latency, and network traffic.

In summary, SCDR disseminates satellite data in a timely and efficient manner to STAR users, especially the calibration and validation and science monitoring teams, helps them generate better products and services for weather prediction, atmosphere, ocean and land surface monitoring, and efficiently facilitates primary research and development activities of STAR.

System Architecture



Multiple Query Interfaces

● Perl Utility Named *scdr-files*

- ❖ Be executed with options like Unix command
- ❖ Locate files of interest with full path directly
- ❖ Be combined with user's task in cron job
- ❖ Offer detailed help information and examples

```
NAME
  scdr-files - Easily locate files stored in the STAR Central Data
  Repository
SYNOPSIS
  scdr-files [OPTION...] DATE...
DESCRIPTION
  scdr-files is a command-line program for locating files stored in the STAR
  Central Data Repository (SCDR). It is designed as a typical UNIX utility
  -- to perform a specific task reliably and fast, files which satisfy the
  search conditions are printed to stdout with full NFS paths, one per line,
  sorted by file's data start time. Any warning or error message is printed
  on stderr.
  Typical usage would consist of invoking scdr-files with search criteria
  (see "OPTIONS" and "EXAMPLES"), then reading stdout and checking the
  program's exit status. Users who cannot pipe stdout into their software
  can redirect it to a temporary file and then read it in. A non-zero exit
  status indicates that no output should be expected either due to an
  execution error or no files found.
```

● SCDR Web Portal

- ❖ Provide an intuitive and interactive Web interface
- ❖ Support selection of start/end observation date and time, ingested date and time, date gap, data type, and platform
- ❖ Make near-real time operational repository information accessible
- ❖ AJAX powered Web application based on open source framework

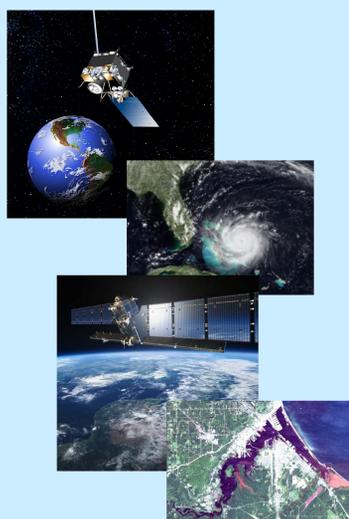
● Lightweight RESTful Web Service

- ❖ Support multiple programming languages (e.g. *Perl*, *PHP*, and *Python*)
- ❖ Be utilized by command-line tools or programs across operating systems
- ❖ Offer a flexible alternative for returning files
- ❖ Provide multiple output formats (*text/plain*, *text/html*, *application/xml*, and *application/json*)

Near Real-Time Satellite Data

SCDR collects over than 200 datasets (about 29,000,000 files, 330+ TB) in multiple heterogeneous formats from:

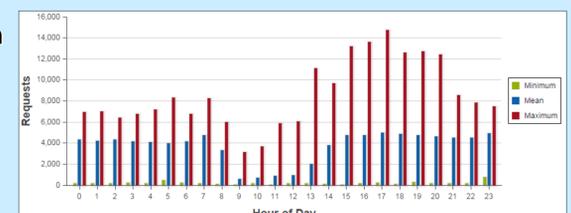
- NOAA Series (6 – 19)
- GOES Series (10 – 15)
- Aqua MODIS
- DMSP F Series (13 – 19)
- Metop-A/B
- WINDSAT
- Suomi-NPP
- Sentinel-1A
- Himawari-8
- More... (in future)



Note: Current SCDR capacity in house is 880TB.

Request Statistics

- ❖ Display detailed SCDR Usage information
- ❖ Provide interactive tables and graphs
- ❖ Create periodical reports in seconds



Name	Description	Stored Files	Stored Volume	Requests	Files	Volume
ghm100	VIIRS Moderate Resolution Band 05 SCDR in HDF5 format	122,192	3.0 TB	671,580	6,235,860,260	154,836.9 TB
sv005	VIIRS Moderate Resolution Band 05 SCDR in HDF5 format	122,213	599.6 GB	654,025	421,205,101	2,012.2 TB
sv007	VIIRS Moderate Resolution Band 07 SCDR in HDF5 format	122,219	927.4 GB	653,956	421,205,380	1,875.7 TB
sv010	VIIRS Moderate Resolution Band 10 SCDR in HDF5 format	122,285	462.3 GB	653,952	421,202,951	1,353.3 TB
sv011	VIIRS Moderate Resolution Band 11 SCDR in HDF5 format	122,213	364.0 GB	653,884	421,202,978	1,223.4 TB
sv009	VIIRS Moderate Resolution Band 09 SCDR in HDF5 format	122,248	582.3 GB	653,847	421,201,052	1,952.7 TB
sv008	VIIRS Moderate Resolution Band 08 SCDR in HDF5 format	122,301	473.0 GB	653,754	421,198,733	1,402.8 TB

Host	Requests	Files	Volume	User	Requests	Files	Volume
rhe01139	5,491,435	3,774,013,525	31,267.1 TB	user1	6,550,626	4,407,816,541	37,559.1 TB
rhe01133	1,058,165	753,688,640	6,291.8 TB	user1	518,509	562,529,401	13,956.8 TB
rhe01045	531,141	5,772,187,185	143,263.3 TB	user2	327,751	5,354,624,978	130,737.6 TB
rhe01073	260,442	103,412,834	427.0 TB	user2	201,467	3,534,000	56.5 TB
rhe01075	126,829	1,731,703	27.3 TB	user2	77,941	1,060,127	13.0 TB
rhe01103	70,095	3,241,113	16.4 TB	user5	71,885	68,446,333	454.8 TB
rhe01048	70,095	3,241,113	16.4 TB	user5	71,885	68,446,333	454.8 TB