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 It is also responsible for managing disk space and removing the outdated files.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
- 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