7.2 ACCESSING NPP DATA FROM CLASS, A TUTORIAL

Drew Saunders*, STG Inc, Asheville, NC
Nancy Ritchey, National Climatic Data Center (NCDC), Asheville, NC
Robert Rank, Office of System development (OSD) Comprehensive Large-data Array Stewardship System (CLASS) Development

1. ABSTRACT

The National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP) and Joint Polar Satellite System (JPSS) operational data products will provide global coverage of environmental conditions associated with the Earth's weather, atmosphere, oceans, land, and near-space environment. This information represents a tremendous increase in data volume compared to that from the existing heritage systems. This increase in volume will provide many challenges for the successful archive and stewardship of this important environmental data resource.

The National Oceanic and Atmospheric Administration (NOAA) National Data Centers (NNDC) are responsible for the scientific data stewardship of the products and information available from the NPP and JPSS missions. The scientific data stewardship consists of long-term archive and preservation, enhancement of the information, and providing easy and efficient access for users. The NNDC consist of the National Climatic Data Center (NCDC), National Geophysical Data Center (NGDC), and the National Oceanographic Data Center (NODC). Currently, the NNDC archive and serve data from many existing missions, including Geostationary Operational Environmental Satellites (GOES) and Polar Operational Environmental Satellite (POES).

This presentation will illustrate how to access the NPP data through the NNDC Comprehensive Large Array-data Stewardship System (CLASS). CLASS provides access to the NPP data through the existing and familiar CLASS web interface. Additional search and access capabilities are planned through the CLASS system evolution.

* *Corresponding author address:* Drew Saunders, NOAA National Climatic Data Center, 151 Patton Ave, Asheville, NC 28801; e-mail: drew.saunders@noaa.gov

2. NPP AND JPSS DATA

NPP and JPSS will provide high resolution environmental products for the weather and climate communities. The NPP mission provides continuity of environmental observations for weather forecasts and climate studies. Sensors on board the NPP satellite include:

- The Advanced Technology Microwave Sounder (ATMS) and the Cross-Track Infrared Sounder (CrIS) provides atmospheric profiles of temperature, moisture, and pressure.
- The Ozone Mapping and Profiler Suite (OMPS) collects total column and vertical profile ozone data.
- The Visible/Infrared Imager Radiometer Suite (VIIRS) provides high-resolution imagery and radiometry data.
- The Clouds and the Earth's Radiant Energy System (CERES) provides solar reflected and Earth emitted radiation information.

The environmental information provided by the mission sensors is processed by the Integrated Data Processor Segment (IDPS) to generate operational products that are disseminated to operational users and the NNDC for archive. The ancillary and auxiliary information used in the production of the operational products is archived. With this information products can be studied and reprocessed for further understanding of the environmental measurements.

2.1 Operational Data

The operational Environmental Data Records (EDRs) include products covering ozone, atmospheric, clouds, earth radiation budget, land/water surfaces, sea surface temperature, and ocean color. Each product resulting from the completion of a stage of processing for the operational EDRs are archived. These products

include Raw Data Records (RDRs), Sensor Data Records (SDRs), Temperature Data Records (TDRs), some Intermediate Products (IPs), Application Related Products (ARPs), and EDRs.

IDPS delays for six hours before delivering the operational products to CLASS. This ensures that any late arriving data are included in the CLASS delivery and that the best data available are archived.

Table 1 NPP Data for Archive

Data Type	Data
Operational	Raw Data Records (RDR) Sensor Data Records (SDR) Temperature Data Records (TDR) Intermediate Products (IP) Application Related Product (ARP) Environmental Data Records (EDR)
Ancillary Data	Official Static Ancillary Data Official Dynamic Ancillary Data
Auxiliary Data	Release Packages Software Supporting Data Documentation Test Data Two Line Elements (TLE) Mission Schedules Mission Notices Revolution Number CERES Auxiliary Data
Cal/Val Data	Retained Intermediate Products (RIPs) Cal/Val Study Results

2.2 Ancillary Data

Ancillary Data are data acquired from external providers and used by the IDPS in the generation of the operational products. “Official ancillary data is defined as the designated ancillary data that is used to produce the official set of NPOESS data products that are delivered to CLASS” (CDFCB-X Vol VI). Official Dynamic Ancillary Data (ODAD) are ancillary data that is updated frequently (potentially daily) for the production of SDRs and EDRs. Official Static Ancillary Data (OSAD) are used for processing and are infrequently updated.

2.3 Auxiliary Data

Auxiliary Data are data not included in the operational products that were required by the NPP or JPSS Programs in the execution of the mission or artifacts of processing. Included in the Auxiliary Data are the Two Line Element (TLE) sets, Mission Schedules, Mission Notices, Revolution Number file, CERES Auxiliary Data, and Release Packages. TLE sets can be used to generate position and velocity for the satellite. Mission Schedules contain events and activities for the mission. Mission Notices provide details of upcoming events such as outages, orbital events, maneuvers, launches, and transition to different mission modes. The Revolution Number file identifies the times of ascending node equator crossing. The CERES Auxiliary Data consists of the Application Packet raw telemetry files and the solar ephemeris files.

There are four types of Release Packages available that will enable users to fully understand the processing performed for NPP and JPSS. The Software Release Package contains the complete source code releases or updates for the operational algorithms. This software can be rebuilt and, with the proper execution environment, executed by members of the user community. The Test Data Release Package consists of files needed to test and validate the software builds. The Supporting Data release Package contains processing coefficients required by the operational algorithms. The Documentation Release Package contains system and product documentation that describes algorithms, processing, and the format and content of data products.

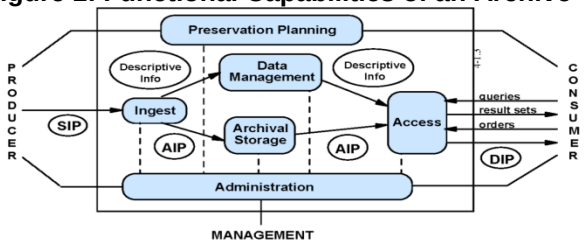
2.4 Calibration/Validation Data

The Calibration and Validation (Cal/Val) Data do not consist of operational products but intend to provide a cal/val record consisting of important data products and studies for understanding the operational products and their evolution. Retained Intermediate Products (RIPs) are non-operational products used in the production of SDRs and EDRs. The Cal/Val Study Results provide findings from important studies.

3. NOAA ARCHIVES

As a part of NOAA’s National Environmental Satellite Data, and Information Service (NESDIS) the NNDC acquire and manage much of the nation’s operational environmental data. As archives, the data centers implement the Open Archival Information System (OAIS) Reference Model defined by the Consultative Committee for Space Data Systems (CSDS) (CCSDS, 2004).

Figure 1: Functional Capabilities of an Archive



The OAIS Reference Model identifies six functional capabilities for an archive. The capabilities include:

- **Ingest** – accept data from producers and prepare the data and information contents for storage and management.
- **Archival Storage** – store, maintain, and provide services for retrieval of the data and information.
- **Data Management** – includes populating, maintaining, and accessing information which identifies and documents archived data and aids in the administration of the data.
- **Administration** – functions required for the overall operations of the archive.
- **Preservation Planning** – monitors the OAIS environment to ensure data availability to the user community over the long-term.
- **Access** – services that support users in searching and obtaining information from the archive.

"Scientific Data Stewardship is a subset of data management, emphasizing data quality, quantification of uncertainty, and long-term data access" (Barkstrom, 2007). Ensuring long-term data access and preservation of information means more than just the progression of 'bits' through technology migrations. It means identifying the content and context of the information contained in the bits and ensuring that information is understood and retained through the migrations. Enhancement of the information entails quality assessments of the measurements and products.

The metadata provided for the operational products are specific to the NPP and JPSS missions. Collection level versions of the metadata have been created adhering to the Federal Geographic Data Committee (FGDC) standards. These FGDC records are stored in the NOAA Metadata Manager Repository (NMMR) which contains metadata for data archived at the NNDC. Those metadata can be harvested and searched using standard tools making NPP and JPSS data accessible to a larger user community.

NCDC, under the Climate Data Records Program, is working to simplify access to the raw measurements. To access raw measurements in the operational RDR a user must understand the application packet bit stream and uncompress any portions that were compressed. This requires specific knowledge of the operational algorithms and software. By reformatting the raw measurements into the netCDF4 data format, the self-describing capabilities of the format can be exploited. With this new record, called a Climate RDR, users will be able to directly access the raw measurements using standard HDF5 and netCDF4 libraries.

The operational products produced by the IDPS are ingested by the NNDC for long-term archive. The IT infrastructure responsible for ingest, storage, search, and dissemination of the archive data is implemented by CLASS. CLASS has provided much of the NNDC infrastructure for years and currently holds approximately 35 different data families including GOES and POES. The number of data families serviced by CLASS is growing rapidly.

The NPP operational data will add approximately 1.4 petabytes (PB) of data annually to CLASS. The NPP data volume is in addition to an existing ingest rate of 2.4 PB in FY10. In the next few years, with the addition of NPP, JPSS, and GOES-R satellites, the volume of data ingested and disseminated by CLASS will grow exponentially. Providing efficient and reliable access to these data is a challenging endeavor. This challenge must be met by utilizing existing tools and services and providing required enhancements to meet future needs.

4. NPP DATA ACCESS VIA CLASS

This tutorial focuses on accessing NPP data via the CLASS web interface. Currently, the web interface is the primary interface but CLASS is undergoing a system evolution that is enhancing existing capabilities and will add new interfaces.

4.1 Search and Display

Through the CLASS web interface common search utilities are provided. A normal search

allows users to search based on satellite, sensor, product type, temporal or geo-graphical attributes. A dataset name search allows the user to identify a specific filename or filename pattern. There is also a collection level metadata search that utilizes the NMMR metadata to search across all data families, including NPP and JPSS.

New searches have been added that are NPP-specific. A granule ID search utilizes the provenance metadata and returns all of the products used in the production of a specific product. For instance, if the granule ID of an EDR is entered, the RDR and SDR inputs to the EDR's processing are returned. A search based on orbit number returns all products produced during a specific orbit.

Each of the searches returns the "result set" containing the products matching the search criteria. For each product in the result set detailed granule level metadata are displayed. After browsing through this information a user selects the products desired for order. A simple search option is available if the user does not require this browse capability. Using simple search the result set is copied directly to an order form.

4.2 Order

The CLASS web interface provides two mechanisms for ordering. Users can order by selecting products from the result set of a search and have the products copied to the order form. On the order form, users select additional products to be added to the order. These may include geo-location products to be packaged with the data products or the inclusion of specific ancillary or auxiliary information. For the web interface, there is a limit on the number of files per order to ensure CLASS can continue to provide the expected level of service. With the increased volume of NPP and JPSS products these limits will have to be tuned to achieve maximum performance. If necessary, users can and should continue to request bulk orders through the CLASS Help Desk.

CLASS provides a subscription option for users interested in automatic dissemination of specific products. The requested products are continually 'pushed' to the user until the subscription is canceled.

4.3 Dissemination

Currently, CLASS supports two methods of data delivery: File Transfer Protocol (FTP) and bulk delivery. Using FTP, users can 'get' their ordered files from a designated holding zone. For larger orders, bulk delivery is required. These orders are processed manually and disseminated using FTP or magnetic disk. Work is underway to automate the bulk delivery process.

4.4 Future Enhancements

CLASS is currently undergoing a system evolution that includes enhancements to many of its core functionalities. These enhancements will make CLASS a more flexible and efficient system to meet the future demands of these expanding data families. Some of the most intriguing enhancements are new access mechanisms for the CLASS data stores. New Application Programmer Interfaces (APIs) are being developed to provide automated access to the data, providing a powerful alternative to the current web interface. Also planned, is the integration of existing interfaces with the CLASS data stores. For instance, Climate Forecast System Reanalysis (CFSR) data stored in CLASS are now accessible through the NOAA National Operational Model Archive & Distribution System (NOMADS) interface.

5. SUMMARY

The NPP and JPSS missions will provide important environmental measurements that are required for current and future scientific research. Protecting these information assets and making them easily accessible is the responsibility of the NNDC. How to utilize CLASS to access that data was covered in this tutorial.

6. ACKNOWLEDGEMENTS

Alexandra Borleis (CLASS) and Stephen Milinovich (CLASS) provided access and screen snapshots from the CLASS NPP test environment at (NSOF). These snapshots slides make up the bulk of the "Accessing NPP Data from CLASS" tutorial.

7. REFERENCES

Barkstrom, B.R., Bates, J. J., and Privette, J. 2007: NOAA's Scientific Data Stewardship Program: Connecting Essential Climate Variables to Science Questions and Societal Impacts for Long-term Information Access, *11th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS),* 87th Annual Amer. Meteor. Soc.

Mtg., J4.6 [Available online at http://ams.confex.com/ams/pdfpapers/117342.pdf]

Consultative Committee for Space Data Systems, 2002: Reference Model for an Open Archival Information System (OAIS), [Available online at http://public.ccsds.org/publications/archive/650x0b1.pdf]

NPOESS Common Data Format Control Book – External, Volume VI – Ancillary Data, Auxiliary Data, Messages, & Reports, No. D34862-06, Rev. H [Available online at http://jointmission.gsfc.nasa.gov/project/sciencedocuments/072010/D34862-06_H_CDFCB-X_Volume_VI_ECR_1016A.pdf]