

# **CLASS PLANS PROVIDING ARCHIVE, ACCESS, AND DISTRIBUTION SERVICES FOR GOES-R DATA**

Robert Rank<sup>1</sup>, Scott McCormick<sup>2</sup>, Constantino Cremidis<sup>3</sup>

<sup>1</sup>NOAA/NESDIS/OSD/GSD, USA, <sup>2</sup> Global Science & Technology, Inc., USA, <sup>3</sup>CSC, USA

## **Problem:**

The Comprehensive Large Array-data Stewardship System (CLASS), a National Oceanic and Atmospheric Administration (NOAA) IT enterprise solution, supports NOAA's data archive and science data stewardship missions by providing the IT portion of an archive for environmental data. CLASS requirements are defined by NOAA's data centers in conjunction with their users and with data producers to ensure that science and other user requirements are clearly defined with respect to NOAA's archive. Currently, CLASS is part of a major 10-year growth program to add new data sets and functionality to support a broader user base. NOAA is defining the types of data to be archived, metadata managing standards, and the data search, display, and delivery services that CLASS will provide to users. These requirements are captured in documents that include requirements specifications, Interface Control Documents, and Submission Agreements that drive the software and hardware architecture changes needed to handle the expected future increases in user and data volumes. The paper will present CLASS' approach for other major data campaigns such as NPP, NPOESS, and NEXRAD, historical data, and its plans for going forward with the GOES-R data campaign.

## **Methodology:**

Placing data on-line for access via the Web is a high priority in accordance with the Federal Government's eGov initiative. Data storage and retrieval systems will continue to be upgraded to support effective and efficient access with special focus on Internet interfaces, and on-line data that support the objectives of the CLASS concept of operations and ensure that the Nation has access (including Section 508 compliance) to their data and information.

New satellite observation campaigns are being prepared for launch and operations. The volumes of data to be collected by these campaigns dwarf the data streams managed by existing archive and distribution systems within NESDIS. The size, number, and frequency of data sets to be stored and distributed will require significant expansion of capacity for moving, storing, processing, and distributing data. New and continuing remote-sensing campaigns include GOES, POES, DMSP, NEXRAD, NPP, NPOESS, Jason-2/3, and METOP. Numerous in situ observation programs also contribute to the information processing challenge.

CLASS has also been designated as NOAA's IT enterprise solution that will support the long-term archive requirements for NOAA data and products. The CLASS concept was developed as a framework to provide integrated data support while accomplishing the needed capacity expansion.

#### *CLASS Approach for New Data*

Considering the importance and complexity of early identification of metadata requirements in particular and the problems associated with data management in general, the CLASS project has a team dedicated to work the details of adding new data into CLASS: the CLASS Requirements Definition Team. The CLASS Requirements Definition Team serves as the technical liaison between CLASS and the NOAA data centers and external stakeholder organizations during the process of negotiating and defining data management requirements for potential, new data campaigns. The CLASS Requirements Definition Team also performs gap analysis and requirements management for these campaigns.

#### *CLASS Experience with New Campaigns*

The CLASS Requirements Definition Team is currently engaged in two large new campaigns and has recently completed two others. The two current campaigns are the National Polar-orbiting Operational Satellite System (NPOESS) and the Geostationary Operational Environmental Satellite (R-series). The campaign recently completed are the Meteorological Operational satellite programme (MetOp) a joint venture between the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and NOAA, and Jason-2. In all cases, CLASS has followed or is following the Open Archival Information System – Reference Model (OAIS-RM). While working with these new campaigns, CLASS has learned a variety of

lessons and identified various opportunities for improvement. The two most important opportunities for improvement identified by the CLASS Requirements Definition Team are:

**First**, although the OAIS-RM indicates that a Submission Agreement is necessary between the archive and the producer, CLASS' experience with these three campaigns show that two additional documents are, in some occasions, necessary to complement the Submission Agreement. These two documents are: an Interface Control Document (ICD) documenting the technical aspects of the electronic data transfer i.e. defining network channels, file and data formats, processing workflow, and acknowledgment protocol. And an Operations Agreement (OA) to document the operational aspects of data transfers i.e. IP addresses and routine and emergency point of contacts. This is due to the fact that for most very large data campaigns, the producer of the data is a science organization that it is concerned with the science aspects of producing the data, while a different group is concerned with the operational aspects of timely creation and distribution of the data. Although both science and operational organizations could be under the same parent organization, CLASS' experience is that there are enough differences to warrant the development of three different, albeit related, documents

**Second**, during the early development of new campaigns the User Community and the Archive, represented by the NOAA National Data Centers (NNDCs) and CLASS, need to actively engage the data producer to define standards for data formats and metadata, and to produce the necessary documentation that must be made available once the data is produced. The development of large data campaigns (like GOES-R) present a unique opportunity to the user community to steer the data producer towards producing data in formats that are easy to work with, that are compatible with the tools used by the user community and that has enough documentation to enable continued usage of the data even after the 'experts' have move to new projects.

The Archive's interests are similar to those of the user community, but for different reasons. In regard to data formats, the Archive is interested on data formats that (1) are directly usable by the user community, therefore do not require the Archive to make data format conversions before distributing the data; (2) facilitates data preservation by not relying on specific software to read it for processing it and; (3) facilitates archive, by being compact and standard. A compact file is easier to receive, copy, archive, and distribute; a standard data format is easier to process as most likely there are already tools capable of processing such format and it is also likely that the

Archive has already experience with such data format. In regard to metadata and documentation, the Archive is interested in the availability of collection and granule level metadata that facilitate data discovery, and enough documentation to ensure long-term information preservation.

### **Central Conclusions:**

The significant increase in high-quality digital data delivered by GOES-R, along with other large-volume data contributors such as NPP/NPOESS and NEXRAD, have led NOAA to develop an agency-wide solution for large-array data archive and access. This new system, the Comprehensive Large-Array data Stewardship System (CLASS), is currently being implemented by the Office of Systems Development (OSD) and the three NOAA National Data Centers. The existing deployment of CLASS has operational nodes at the National Climatic Data Center (NCDC) in Asheville, NC, and the National Geophysical Data Center (NGDC) in Boulder, CO. The centralized development/integration/test environments are retained at OSD in the new NOAA Satellite Operations Facility (NSOF) in Suitland, MD. The new CLASS operational node at NGDC will have the advantage of being accessible by the extensive research community in the Boulder region via broadband connection. NGDC recently implemented the National Lambda Rail to allow broad band data access via the commercial internet. The NSOF development/integration/test environment enhances for sharing resources among development and integration environments.

Effective systems must be in place to support the scientists and research community, as well as address data management issues that are associated with massive volumes of data. Therefore, CLASS will implement an architecture for an integrated, national environmental data access and archive system to support ingest, archive, access, and long-term stewardship requirements of the large array, remotely sensed data sets.

### **Bibliography:**

CLASS Level 1 Requirements Document and Mission Success Criteria, Version 1.16, Nov 2008

National Oceanic and Atmospheric Administration Enterprise Architecture Version 1.61

CLASS Target System Architecture Overview Initial Release March 2009

Open Archival Information System – Reference Model (OAIS-RM), ISO standard