P1.20

AVAILABILITY OF GEOSPATIALLY ENABLED NOAA SATELLITE PRODUCTS

David J. Vercelli*[1]
National Environmental Satellite, Data, and Information Service (NESDIS)
National Oceanic and Atmospheric Administration (NOAA)
Silver Spring, Maryland

Co-Author: Howard J. Diamond (NOAA/NESDIS)

1. INTRODUCTION

The National Oceanic and Atmospheric Administration (NOAA) is the federal agency with statutory responsibility for long-term archiving of the Nation's environmental data. NOAA spends almost a billion dollars each year collecting environmental data from around the world and from space in support of its mission. NOAA's vast data holdings are collected and stored in many different facilities across the country, some of which are responsible for the perpetual stewardship, archiving, and dissemination of environmental data. While in the past, real-time access to environmental data was required only of NOAA's National Weather Service (NWS) to support weather warnings and forecasts, today there are demands on virtually all of NOAA's programs to provide information on the health of the environment in real-time. For the 21st century, NOAA envisions its data and information products to be available to the Nation as part of a national decision support system for the purpose of the following:

- Saving lives and protecting property
- Promulgating public policy
- Managing and conserving living marine resources
- Enhancing the economic prosperity and quality of life in the United States

Within NOAA, the National Environmental Satellite, Data, and Information Service (NESDIS) is responsible, among other things, for the collection, archiving, and dissemination of environmental data collected by a variety of *in situ* and remote sensing observing systems from throughout NOAA and from a number of its partners such as the National Aeronautics and Space Administration (NASA). NESDIS has been collecting these data for more than fifty years and will be archiving and disseminating even larger data sets in the future. In order to prepare for this large increase in its data holdings, NESDIS initiated the planning, development, and implementation of the Comprehensive Large Array-data Stewardship System (CLASS) that will provide long term archive and access services for these data.

CLASS is being designed to accommodate the data flow from current satellite-based (e.g., Polar-orbiting Operational Environmental Satellites - POES, Geostationary Operational Environmental Satellites - GOES, Defense Meteorological Satellite Program - DMSP) and ground-based (e.g., Next Generation Weather Radar - NEXRAD) observing systems. It also is being structured to handle the large increases in data that will come from planned satellite launches [e.g., European Meteorological Operational satellites (MetOp), National Polar-orbiting Operational Environmental Satellite System (NPOESS), NPOESS Preparatory Project (NPP), and Earth Observing System (EOS) satellites]. Finally, it must be capable of supporting current *in situ* data sources (e.g., Automated Surface Observing System - ASOS).

2. NESDIS MISSION

The NOAA NESDIS mission is to provide and ensure timely access to global environmental data from satellites and other sources to promote, protect, and enhance the Nation's economy, security, environment, and quality of life. To fulfill its responsibilities, NESDIS acquires and manages the Nation's operational environmental satellites, provides data and information services, and conducts related research (NESDIS, 2002).

Within the NESDIS structure, a large portion of the Nation's current archive of environmental data is stored and maintained by NOAA's National Data Centers (NNDC). The NNDC consists of the National Climatic

* *Corresponding author address*: David J. Vercelli
NOAA/NESDIS, 1335 East-West Highway, Room 8126, Silver Spring, MD 20910;
e-mail: david.vercelli@noaa.gov

Data Center (NCDC) in Asheville, NC; National Oceanographic Data Center (NODC) in Silver Spring, MD, including the National Coastal Data Development Center (NCDDC) in Bay St. Louis, MS; and National Geophysical Data Center (NGDC) in Boulder, CO. In addition to the NNDC, the Satellite Active Archive (SAA) in Suitland, MD provides electronic distribution of data and derived data products from the POES satellites.

The mission of the NNDC is to provide data management in the broadest sense. The NNDC plays an integral role in the nation's research into the environment, and at the same time provides public domain data to a wide group of users. This user community includes:

- private industry
- universities and other educational facilities
- research organizations
- federal, state, and local governments
- foreign governments, industry, and academia
- publishers and other mass media
- the general public

3. BACKGROUND

No previous decade has seen the magnitude of changes in the volume of data coming into NOAA for processing and archive as those experienced in the 1990s. Already, there are significant new volumes of data from NEXRAD and DMSP being preserved as part of the NOAA archives. However, that explosive growth is nothing compared to what is planned between now and 2015 (Figure 1). Even as current observing systems continue to provide data, new satellite systems such as MetOp, EOS, and NPP will be going into operations within the next few years. In addition, NPOESS and the next generation GOES (GOES-R) will follow towards the end of this decade. These systems will provide orders of magnitude more data which will present formidable challenges for NOAA. At the same time, new *in situ* observations from widely dispersed automatic reporting platforms are generating significant increases in conventional observation data which NOAA will be required to manage for the long term.

In general, user requests for NOAA environmental data and information increased throughout the 1990s. However, with the growth of the World Wide Web as a ubiquitous technology, a global market was created nearly overnight for NOAA's data and information services. And, although off-line data requests doubled, the truly exponential growth has been in the number of on-line users who extend far beyond NOAA's traditional user community. While these on-line requests have dramatically increased, it is important to realize that only a portion of NOAA's data archive currently is available to the user on-line. As on-line access to NOAA's data expands, the user's average level of technical sophistication and scientific expertise is changing. On-line users now want to search for information and answers to specific questions rather than simply for access to data. Users, no longer content to wait days for their data or information, are

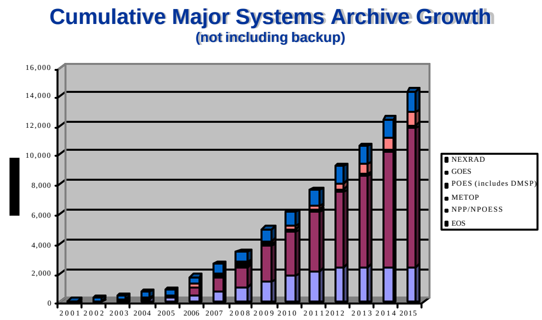

Figure 1

demanding on-line ordering, search, and browse capabilities with electronic file transfer for data delivery. New user groups require near-real-time access to data to support decision-making and rapid response needs. Increasingly, users want information rather than data, as information and products derived from observations are frequently more useful to business and industry than the original data. Scientists and advisors have a critical need for long time-series of historical and recent environmental data to assess long-term trends, evaluate current status, and predict future conditions and events. Therefore, the timeliness and completeness of NOAA's environmental records are crucial.

4. CLASS

Now through the judicious application of recent technological advances, it is possible to provide incremental improvements in service capabilities at reasonable costs. In an effort to take full advantage of these opportunities, NESDIS has initiated an environmental data and information archiving and access activity which will focus on improving the ability of NESDIS to be the steward for NOAA's environmental data and information and to maintain a permanent archive that is easily accessible to the world science community and to other users. The keystone of this activity is the Comprehensive Large Array-data Stewardship System (CLASS) Project.

The CLASS Project is designed to enhance the NOAA capability to provide environmental data and information archive and access services to the Nation through the effective application of modern, proven techniques and technology. The project places special emphasis on the ability to efficiently archive the vast quantities of NOAA satellite and *in situ* observational data currently being collected and to be collected; to safely and permanently preserve those valuable data for future generations to use; and to provide rapid data access in a cost-effective and secure manner.

There is reason to expect that the information technology advances we have seen in the last ten years will continue for the foreseeable future. With these advances, NOAA has made significant progress in its ability to archive and provide access to its increasing data volumes, and will continue to leverage on these advancing technologies. Management of these data can be accomplished only through a rapid expansion in storage capacity, increased communications bandwidth, and automation of the means of data ingest, quality control, and access. The CLASS project will act as the connection in NOAA's effort to meet these challenges and pave the way to accommodate the additional massive data volumes expected over the next several years.

There are a number of aspects to the successful implementation of CLASS. First, there are those aspects which are purely mechanical or technical in nature. They include, for example, communicating the data from the source to the primary and backup storage locations; quality control and pre-processing of the data; storage of the data on media such as tapes and disk; and, post-processing the data to extract information. In addition, there are those issues concerning the virtual on-line search, discovery, retrieval, display, and customer order processing capabilities for the user community. All these various tasks must be accomplished securely, quickly, and efficiently to meet the needs of NOAA's user community.

The ability to ensure on-going scientific stewardship for NOAA's environmental data and information will only be possible through extensive enhancement of NOAA's current data ingest, quality assurance, storage, retrieval, access, and migration capabilities. This goal will be met through the development and implementation of a uniform archive management system, which will be integrated with a large-volume, rapid-access storage and retrieval system capable of storing the incoming large array environmental data, *in situ* data, and operational products as well as receiving a user's on-line data request, automatically processing the request, and providing the requested data on an appropriate media. This system will provide standardization in media, interfaces, formats, and processes. Additionally, the system will facilitate ongoing migration, preservation, and validation to new technology and media. This system is modular in design, built to integrate with automated real-time or near-real-time systems that deliver data. Transaction processing will be implemented to enable essentially autonomous operations and, where appropriate, the system will allow users to pay for data or services through credit card or automated billing.

The target architecture goal will, through life cycle replacements and upgrades, bring the current NOAA National Data Centers under a single archive and access architecture that will be under formal configuration management control. This will allow elimination of duplication of effort, minimize stand-alone systems, build the infrastructure to accommodate the large array data sets, and reduce the overall operational and system maintenance costs. The foundation system that is being used for CLASS is the highly successful and stable SAA to provide the maximum flexibility while minimizing development work and costs.

5. GEOSPATIAL ENHANCEMENTS TO CLASS

CLASS is being designed to allow improved access to the wealth of satellite data via the latest data discovery and geospatial techniques and tools. This effort is being led by the CLASS team primarily located at NGDC with assistance from NODC and NESDIS headquarters.

A critical aspect to making these data readily available to our customers is to ensure that the satellite metadata are current, accessible, and that they follow the appropriate metadata standards (Habermann, et al, 2002). Currently, the NESDIS Satellite Metadata is physically located at several different locations and only minimally takes advantage of national (Federal Geographic Data Committee, FGDC) and International Standards Organization (ISO) draft ISO 19115 metadata standards. As a note, the current FGDC standard will be ISO 19115 compliant. To correct this situation, the CLASS project has initiated a number of tasks to improve our current data discovery and geospatial techniques and tools, and to develop new ones as appropriate.

One task is designed to implement a robust NOAA Metadata Repository (NMR) management system within CLASS. Metadata, which describe the content, quality, condition, and other characteristics of data are critical to data discovery, usability, quality, interoperability and automatic processing. The planned NMR architecture will primarily handle large array data sets but also be capable of handling smaller data sets as well. The plan for developing the NMR represents a comprehensive approach to creating high-quality metadata and revitalizing NESDIS systems for managing it. The proposed effort relies on state-of-the-art data management tools and techniques and extensive experience and expertise in the NESDIS Data Centers and satellite data processing groups. The plan emphasizes building on this expertise to create a metadata management and access system that will serve as the preeminent example for other NOAA Line Offices and agencies at all levels of government.

The NMR will be based on relational databases (Oracle), internet mapping software (ArcIMS), eXtensible Markup Language (XML) and XML stylesheets, and Java access tools (Blue Angel Technologies). The NESDIS metadata team has extensive experience with all of these tools and they are presently running in several NESDIS locations as part of the National Virtual Data System (NVDS) and Coral Reef Information System (CoRIS) Projects. This plan for the development of the NMR builds upon the work done by NVDS and CoRIS over the last several years. That work has included populating repositories at NGDC (NVDS) and NODC (CoRIS) with nearly 10,000 metadata records from the Data Centers, other NOAA Line Offices, and CoRIS investigators, and developing and field testing several metadata manager interfaces to the repository. In addition, this project will incorporate work done developing programmatic interfaces to the repositories as part of the overall CLASS Data Discovery Support System.

The foundation of the proposed NMR system will be made up of Oracle databases with access provided using the Enterprise Blue Angel Metadata Repository Java Classes. This foundation also includes a set of XML files that support Z39.50 searches, metadata transfer to other systems [e.g., Global Change Master Directory (GCMD), the FGDC Clearinghouse, and the Geography Network)], and display using XML stylesheets. It will also rely on a spatial database (either Oracle or Informix) and ESRI's geographic information systems (ArcSDE, ArcMAP, and ArcIMS).

Initially, the focus will be on two metadata collections, the SAA Data Family Metadata and the Satellite Data Products Database. These were chosen because these two data collections play a critical role in the creation of FGDC compliant metadata for the existing NESDIS satellite products. The primary goal of this study is to compare the individual fields in the existing metadata collections and to evaluate those fields that have similar meanings (semantics). The FGDC standards with the remote sensing extensions are being used for this crosswalk process. This combination provides a broad collection of well defined metadata fields and a framework for organizing the crosswalks.

The geospatial foundation to be used in CLASS will be based upon standard relational database management systems with spatial extensions (Habermann, 2002). This is being done since these systems are capable of supporting the wide variety of environmental data types to be available via CLASS. In addition, these systems offer support for a number of different data access approaches being investigated.

As such, a second task initiated will work towards the development of tools for processing level 1 and level 2 data to extract the spatial information presently stored in the file record headers and to ingest that data into geospatial databases as orbits and scan lines. Completion of this task will immediately benefit CLASS customers by making all of the data in the database available for both spatial and general SQL queries.

6. CONCLUSIONS

National population and industrial growth, coupled with increased overall public awareness, concern, and dependence on climate data and information, has resulted in an increased demand for more accurate and more timely observations of our National and global environment over longer and longer time periods in efforts to predict our future environment.

Through CLASS, NOAA will continue to be a leader in archiving and providing access to the wealth of environmental data and information available today from existing satellite and ground-based observing systems and in the future as new systems are brought on-line.

Expected performance outcome of the CLASS Project will be greatly enhanced communications capabilities; increased computer storage capabilities; increased computer power, the use of commercially available, modular hardware and software; and improved World Wide Web access to the data and information through new or enhanced database management, search, order, browse, and subsetting techniques. In addition, through the implementation of a strong metadata management system and the expanded use of geospatial tools and techniques, CLASS will provide a seamless data archive portal for the Nation and the world.

7. REFERENCES

Habermann, T., 2002: Geospatial Data Systems and NESDIS Satellite Data (unpublished)

_____, K. Nuttycombe, J. Barkley, T. Gaines, and T. Stevens, 2002: NESDIS Satellite Metadata Model, V1.0 (unpublished).

NESDIS, 2002: http://www.ngdc.noaa.gov/ngdcinfo/aboutngdc.html

8. ACKNOWLEDGMENTS

We acknowledge all of the dedicated employees at each of the NOAA National Data Centers. Without their hard work and commitment, there would be no long term archive of environmental data and information. Specifically, we want to thank Ted Habermann of NGDC, and Parmesh Dwivedi of NODC, whose vision and expertise have helped define the tasks to be accomplished to ensure CLASS has the quality metadata available to it and the tools in place to allow our customers to search, browse, and access the archive data and information. Ted, Parmesh, and their staffs are primarily responsible for the metadata and geospatial efforts within CLASS.