5.1

NOAA's SCIENTIFIC DATA STEWARDSHIP PROGRAM

John J. Bates*
NOAA's National Climatic Data Center, Asheville, North Carolina

1. INTRODUCTION

The importance of understanding and predicting climate variation and change has escalated significantly in the last decade. To integrate federal research on global climate change, President George W. Bush announced in February 2002 the formation of a new management structure, the Climate Change Science Program (CCSP). The National Academies' National Research Council (NRC) has recommended several research priorities for climate research and the CCSP drafted a strategic plan, which was reviewed by the NRC and the CCSP issued a final plan. (USCCSP, 2003)

Climate: Long-term averages of weather conditions. Typically, 30 years of data have been used to form the climatological means.

Climate Variability: Variations from long-term average weather conditions for time periods of a month or more.

Climate Change: Changes in the long-term averages of weather conditions, e.g., global warming, onset of ice ages.

In response to these national initiatives in climate science, the NOAA National Environmental Satellite Data and Information Service (NESDIS) will develop a NOAA Plan for Creating Climate Data Records from Operational Satellites to provide a framework for the use of climate data from existing and new instruments aboard NOAA satellites, including instruments on the National Polar-orbiting Operational Environmental Satellite System (NPOESS). The goal of the plan is to ensure that satellite climate data are processed, archived, and distributed to users in a manner that is scientifically defensible for monitoring, diagnosing, understanding, predicting, modeling, and assessing climate variation and change.

The NOAA mission is to understand and predict changes in the Earth's environment and conserve and manage coastal and marine resources to meet the Nation's economic, social and environmental needs. NOAA has statutory responsibility for long-term archiving of the nation's environmental data and has recently integrated several data management functions into a Scientific Data Stewardship initiative. These functions include careful monitoring of observing system performance for long-term applications, the generation of authoritative long-term records from multiple observing platforms, and the proper archival of and timely access to data and metadata.

NESDIS maintains the world's largest archive of climate-related data and information spanning the ice age to the space age. NESDIS also operates the Nation's operational satellite observing system and is developing and implementing the U.S. Climate Reference Network. Data and information from NOAA space-based and ground-based observing systems are used along with other climate-related NOAA and non-NOAA observing system data to construct long-term records regarding local, regional, national, and global climate variability and change. Within NESDIS are three large data centers that archive and provide access to climate, ocean, and geophysical data. The National Climatic Data Center (NCDC) in Asheville, North Carolina is the largest archive of weather data in the world.

Climate research is generally based on data collected for other purposes, primarily for weather prediction. To make these data useful for climate studies, it is usually necessary to analyze and process the basic observational record to create a *Climate Data Record* (CDR). A CDR is a series of observations over time that

**Corresponding author address:* John J. Bates, NOAA's National Climatic Data Center, Remote Sensing and Applications Division, Asheville, NC 28801-5001; e-mail john.j.bates@ noaa.gov

measure variables believed to be associated with climate variation and change. These changes may be small and occur over long time periods (seasonal, interannual, and decadal to centennial) compared to the short-term changes that are monitored for weather forecasting. Thus, it is usually necessary to construct a CDR from data that span long time scales and sometimes from multiple data sources. Scientists must characterize and quantify the sensor, the spatial and temporal errors of these diverse and frequently large data sets in order to produce a sufficiently accurate time series for studying trends in climate variability and change.

The NOAA/NESDIS operational satellite data program currently collects, receives, produces, distributes, and archives data about climate, including the climate satellite products shown in the box. Many of these data are processed in response to specific requests from the scientific community who need long-time series climate records. In some cases, the raw data and metadata are provided to external investigators such as those in academia, at other U.S. agencies, or those involved in international projects, who produce the climate data records. NESDIS scientists also produce a number of climate products, either in-house or in collaboration with NASA. Examples of climate data records which have been developed from operational satellite observations are shown in Table 1.

Using existing satellite data to produce long-climate records has shown that adapting observations designed for weather prediction to climate issues in an *ad hoc* way is not sufficient to produce reliable findings and to draw reasonable conclusions about climate change. NOAA recognizes that the development of quality climate data records is key and that a program focus on the development, retention, and distribution of climate data records will be necessary to meet the needs of the science community.

Current NOAA Satellite Products

- **Atmosphere**
  - Temperature soundings
  - Moisture soundings
  - Winds
  - Clouds
  - Aerosols
  - Earth Radiation Budget
  - Precipitation
  - Ozone
- **Ocean**
  - Surface temperature
  - Ice cover
  - Surface winds
  - Color
  - Sea level
- **Land**
  - Vegetation condition
  - Snow Cover
  - Surface Temperature

Table 1. Examples of Climate Data Records Based on Operational Satellite Observations

Climate Product	Satellite/Instrument	Produced by	Since
Earth Radiation Budget (ERB) Outgoing long-wave radiation (OLR) Absorbed solar radiation (ASR)	POES/AVHRR	NESDIS	1978
Ozone	POES/SBUV/2 & POES/ATOVS/HIRS	NESDIS/NASA	1985
Blended Sea Surface Temperature (SST)	POES/AVHRR	NESDIS/NWS	1981
DMSP SSM/I Climate Products (rainfall, rain frequency, snow cover, sea ice cover, clouds, water vapor, and oceanic wind speed)	DMSP SSM/I	NESDIS	1987
Vegetation (NDVI and drought index)	POES/AVHRR	NOAA/ORA	1982
Atmospheric Temperature	POES/MSU	Univ.of Alabama	1979

Snow Cover	POES/AVHRR, GOES, Meteosat, GMS Visible imagery, DMSP/SSM/I	Rutgers Univ. Climate Laboratory	1966
Clouds	POES/AVHRR; GOES, Meteosat, and GMS Visible IR imagery	WCRP/Internati-onal Satellite Cloud Climatology Project	1983
Precipitation	POES/AVHRR; GOES; Meteosat and GMS Visible IR imagery,DMSP SSM/I	WCRP/Global Precipitation Climatology Project	1986

Table 1. Examples of Climate Data Records Based on Operational Satellite Observations

Efforts to use the operational environmental satellite observations over the past decade or more have resulted in a set of recommendations from researchers that have recently been formalized by the climate community into satellite climate observing principles (WMO, 2003). Six of those principles are essential topics for discussion and recommendations of this report:

- Development and operational production of priority climate products should be ensured.
- Systems needed to facilitate user access to climate products, metadata and raw data, including key data for delayed-mode analysis, should be established and maintained.
- Continuing use of still-functioning baseline instruments on otherwise de-commissioned satellites should be considered.
- The need for complementary in-situ baseline observations for satellite measurements should be appropriately recognized.
- Network performance monitoring systems to identify both random errors and time-dependent biases in satellite observations should be established.
- Multiple observing and analysis techniques for critical climate data records should be used.

2. CONCEPTUAL FRAMEWORK

For discussion purposes, a framework for accommodating the requirements and lessons learned outlined above and that is consistent with the CCSP strategic plan for monitoring and observing the climate system is presented. The framework has five objectives: 1) develop real-time monitoring of all satellite observing systems, 2) generate CDRs in near real-time, 3) process large volumes of satellite data extending up to decades in length to account for systematic errors and to eliminate artifacts, 4) conduct research by analyzing data sets to uncover climate trends, and 5) provide archives of both raw data records and CDRs, and facilitate distribution of CDRs to the research community. Each phase of this end-to-end system will require collaboration with climate data science teams, input from climate data users, and should leverage knowledge and resources from other climate data programs and organizations. The framework is depicted in Figure 1.

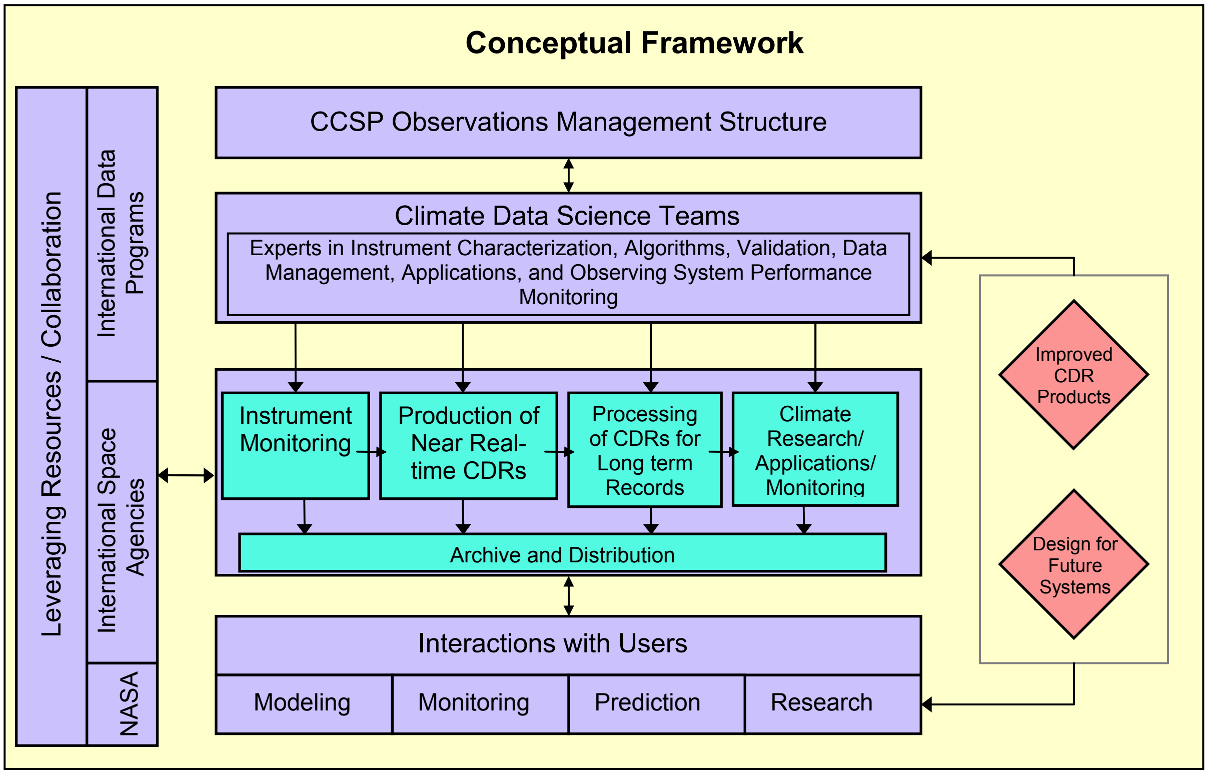

Figure 1. Conceptual Framework for Creating Climate Data Records

2.1 Observing System Performance Monitoring

Understanding the effects of the observing system on the data measurements in real-time will provide data of known quality, and for which temporal and spatial biases can be minimized. Observing factors affecting the data quality include the following:

- Biases inherent in the observing system
- Changes in instruments
- Satellite orbital drift
- System calibration
- Sensor degradation in space
- Satellite to satellite discontinuities
- Satellite or instrument system malfunction

2.2 Near Real-Time CDR Generation

Climate researchers need satellite data on an ongoing basis for prediction of climate variability, such as the El Niño-Southern Oscillation, extreme storm events, excessive rainfall, or drought. Therefore the capability to provide routine observations and generate CDRs in near real-time is needed.

In addition, scientists need to develop data sets that can be compared with or added to the historic record for monitoring long-term climate changes. As data are received, scientists at NOAA will automatically update historical climate data bases to maintain a global climate perspective in near real-time.

Converting the raw data records, sensor data records, and environmental data records (EDR), when applicable, to produce CDR's and science products will be a major challenge that will require leveraging the knowledge of NOAA and NASA researchers, and other members of the scientific community to develop the algorithms. These efforts will involve the following:

- Calibration, inter-calibration and characterization of satellite instruments
- Development of processing algorithms
- Detection and elimination of systematic errors in the data set
- Generation of stable climatic time series
- Validation of data products
- Analysis of data

2.3 Processing

Periodic processing of data sets will be necessary to incorporate new information, new instruments, and improved algorithms.

Periodic processing of the long-term data record may be called for when:

- *An improved algorithm is developed*
  Data will be reprocessed to accommodate the latest scientific findings into the data products. As scientific research improves our understanding of the earth's physical processes, existing algorithms will be refined or replaced with new algorithms.
- *New information on the in-flight behavior of an instrument is obtained*
  Recalibration of measurements may be performed as a result of analysis of instrument behavior.
- *An error is discovered in a processing system*
  A coding or other software error may be discovered in the processing system. This type of error may not be detected in an EDR, but analysis and comparison with other data sets may reveal an error at the CDR level.

2.4 Research and Application

Another component of the framework involves a research activity, as opposed to the "housekeeping" responsibility of processing data sets. In addition to developing CDR algorithms, climate researchers, working with the long-term data record will continue to make contributions to climate change and variation research by analyzing data sets to uncover trends. Activities will include the following:

- Development of climate quality algorithms for creating CDRs.
- Analysis of time series to detect trends that may be emerging from the record, and comparing the results to results of other researchers.
- Joint studies with the climate research community to advance the use of satellite data for climate applications.
- Production of periodic assessments for decision makers, other climate researchers, and the public.

2.5 Data Archive and Access

An operational climate data service must ensure that all climate data are preserved and made available to users. In addition to the climate data; metadata; production software source code; documentation on the data, meta-data and data formats; ancillary data; calibration/validation information; and QA information will also be archived. Regular back-up of data and the capability to migrate any or all of this information to new media are also important.

The primary goal of the NESDIS plan for the generation of climate data records will be to support the user. Therefore, a critical objective is to provide free and open sharing and exchange of climate related data and products. Services will include availability of data in near real-time, and access to both raw radiances and NOAA data products. Community consensus algorithms and techniques will be sought to accomplish these goals. Standards will be developed for the data and media format to be supported for data distribution.

3. DISCUSSION ISSUES

To ensure that CDRs are both accessible and useful to climate researchers and decision makers, and to promote science community participation and consensus, NESDIS is seeking advice from the National Academies (NRC,2004) on a number of issues prior to the development of a plan for creating climate data records from NOAA satellites:

- **Science oversight and participation by external community**
  Examples of functions that could be carried out in this way include algorithm development, generation of CDRs (for the smaller data sets), validation campaigns, research with the data sets, and the evaluation of the data in climate models for other applications.

  Issues: What approach should NOAA consider for obtaining science guidance and oversight for its end-to-end CDR program? What scientific criteria should be used to decide whether a function should be performed in-house or externally?
- **NOAA/NASA relationship**
  NOAA and NASA have a history of cooperative activities related to CDR generation, including the NOAA/NASA

Pathfinder program, the joint NOAA and NASA support of the development of data sets for NOAA's Climate Change Data and Detection project, NOAA scientist participation on NASA science teams that produce CDRs, and the NOAA/NASA cooperation on the generation of a long-term ozone data.

Issues: What are some realistic options for engaging NASA in an end-to-end system for generating CDRs from the operational satellites?

- **Obtaining feedback from users**
  It is extremely important that NOAA obtain feedback from the user community on the utility of the CDRs in various climate applications – climate monitoring and diagnosis, seasonal to interannual climate prediction, decadal scale climate modeling, climate research, and other governmental and private sector applications.

  Issues: What are the optimal mechanisms for entraining the user community in the use, test, and evaluation of CDRs and for obtaining feedback?

- **Participation in international data projects**
  Some of the widely used satellite climate data sets have been produced as part of international programs. To the extent that these programs use operational satellite data and are intended to provide a sustained, long-term record, it seems appropriate for NOAA to actively participate in such programs.

  Issues: What should be the role of NOAA in international data projects? What functions should NOAA perform?

- **Algorithm selection**
  For operational programs, algorithms are selected in a variety of ways. NOAA's satellite EDRs, for example are selected via in-house Product Oversight Panels. NESDIS will explore similar or other viable approaches for selecting algorithms.

  Issues: What are some viable approaches for algorithm selection? Once an algorithm is implemented, what procedures should be used to determine when the algorithm should be updated or replaced? How should replacement algorithms be selected?

4. SUMMARY

NOAA's vision is that through the establishment and execution of an end-to-end system for CDRs, more confident conclusions may be drawn regarding climate variability and change, and that this improvement will benefit policy makers and the public at large. By establishing a programmatic framework to properly address issues surrounding climate data records, NOAA will ensure the quality, usefulness, and accessibility of the data for current and future generations.

5. REFERENCES

National Research Council, 2004: Climate data records from environmental satellites: interim report. Washington, D.C., National Academy Press, 116 pp.

U.S. Climate Change Science Program, 2003: *Strategic plan for the U.S. Climate Change Science Program, final report*. Washington, D.C.: U.S. Climate Change Science Program, 202 pp.

World Meteorological Organization, 2003: *The second report on the adequacy of the global observing systems for climate in support of the UNFCCC*. (GCOS 82; WMO/TD no. 1143) Geneva, Switzerland: World Meteorological Organization, Global Climate