# Scientific Data Stewardship

# FY 2007 Information Sheet

The goal of the Scientific Data Stewardship (SDS) Program is to provide high quality Climate Data Records (CDRs) for data from the atmosphere, oceans, and land surface, where the data are identified as essential climate variables within the Global Climate Observing System. The SDS program is intended to produce CDRs routinely on an operational basis, with an initial emphasis on the use of satellite observations that can demonstrate high levels of maturity in scientific and preservation attributes as well as high societal benefit. Initial development of candidate CDRs should already have occurred within other NOAA or agency programs.

## SDS Detailed Description

The Scientific Data Stewardship Program seeks to support the routine, operational production of Climate Data Records (CDRs) for data from the atmosphere, oceans, and land surface. This production requires collaboration between experts in the climate community and experts in data management, and must be informed by user feedback on the accessibility and usability of the produced CDRs. Applicants to this program must demonstrate that proposed CDRs have reached a critical level of maturity in science and data preservation and are of demonstrable societal benefit to be considered for funding under this Program.

Successful applicants will need to demonstrate success in three areas required for operational production of high quality Climate Data Records:

a. **Operational quality assurance of ingested data and routine monitoring of data quality and provenance during production.** In addition to such simple measures as assuring that ingested data is of the expected size and format, applicants will need to identify and use appropriate tools for detecting and diagnosing subtle spatial and temporal biases in the incoming data stream and in the software that transforms that stream into other data products. Applicants will also be expected to demonstrate that they will be able to use the detection and diagnosis to mitigate these biases and to record the detection, diagnosis, and correction workflows so that data users can verify the uncertainties in the data.
b. **Generation of authoritative, long-term records.** Because Climate Data Records are generated from irreplaceable historical observations, applicants must demonstrate approaches that will preserve and enhance the value of the data. Such approaches are expected to involve rigorous data analysis and research that will validate and improve these records, particularly by intercomparing multiple data sources and reprocessing entire data streams. The approaches must provide visibility into the work done and encourage independent reprocessing efforts, including approaches that reprocess entire data streams from the fundamental measurements when scientific advances and expected societal benefit warrant it.

SDS will give precedence to physics-based techniques to fuse data from disparate observing systems, such as direct measurements from ground-based networks and remotely sensed measurements from satellite-borne instruments. The authoritative nature and vitality of the Climate Data Records will be maintained through peer reviews, user recommendations, and independent processing of data that produces alternative versions of essential climate variables.

c. **Generation of CDR context capable of surviving transformative migration.** To ensure data usability, all Climate Data Records need to record the context that allows future users to understand, modify, and use data, even after the original producers have finished their work and moved on to other endeavors. Furthermore, both the data and the context need to survive transformative migrations, as new hardware and software makes older versions obsolete. Applicants to the Scientific Data Stewardship Program will need to provide approaches to dealing with these critical issues for long-term preservation. It is particularly important for participants in this program to ensure complete archival and access capabilities that include data, metadata, validation and production workflows, source code, and documentation, so that the data and context are comprehensive, complete and usable in perpetuity. Open, efficient access to the metadata, data products, and data streams must be ensured, including provision for making the data available in well-documented, cost-effective formats.

Specific Requirements

Proposals should identify:
1. Designated user communities, identifying the NOAA Goal(s) and Programs(s) that will be advanced as a result of a successful proposal;
2. An endorsing NOAA data center identified as the primary recipient of data products, metadata, and tools developed from the effort;
3. Success criteria and metrics to be used to judge that, by the end of the proposal, the designated user community has received the desired benefits and that the data center can make metadata and tools received operational at minimal expense and with minimal operating costs.

Ultimately, the Scientific Data Stewardship Program seeks to understand climate variability and change by asserting leadership for satellite-based Climate Data Record generation, applying new approaches to generate and manage satellite Climate Data Records, developing new community relationships and virtual organizations, and ensuring long-term consistency and continuity for a satellite Climate Data Record generation program. The Scientific Data Stewardship Program seeks to encompass the full range of institutional diversity within the climate community. To that end, this program seeks approaches that allow organizations producing and archiving Climate Data Records to maintain local autonomy within a context that encourages responsible participation in federations that foster increased data sharing, interdisciplinary data understanding, and improved assessments of data quality.

Bruce Barkstrom of the National Climatic Data Center is the technical lead on the SDS project.

In FY 2007, it is anticipated that approximately $800K will be available and eight projects will be supported.