CGMS XXXII NOAA-WP-23 & 24
Prepared by: NOAA
Agenda Item: II.5 & II.6
Discussed in WGII

NOAA REVIEW OF METADATA RECOMMENDATIONS AND REVIEW OF THE GCOS CLIMATE MONITORING PRINCIPLES

The document provides input to discussions in Working Group II at CGMS XXXI with regard to the current work at NOAA on data and products for climate applications.

This paper responds to Actions 31.31 and 31.32.

CONSIDERATIONS FOR LONG TERM STEWARDSHIP OF SATELLITE DATA FOR CLIMATE APPLICATIONS AND A CONSOLIDATED LIST OF METADATA

1 INTRODUCTION

Discussions at CGMS XXX within Working Group II addressed the opportunity and need to establish data sets from operational meteorological satellites suitable for climate applications. In that regard WG II placed the following two actions which are briefly addressed in this paper (also through cross-referencing to other papers):

ACTION 31.31 Satellite operators to report on their plans to consider the metadata recommendations in EUM-WP-22 when re-transcribing their data archives.

ACTION 31.32 All space agencies to provide information specifically addressing the ten GCOS climate monitoring principles related to the space based component of the Global Observing System in a manner similar to EUM-WP-10.

NOAA has supported climate research activities through the World Climate Research Program (WCRP) International Satellite Cloud Climatology Project (ISCCP), Global Precipitation Climatology Project (GPCP), and related activities within the WCRP Global Energy and Water Cycle Experiment (GEWEX) Radiation Panel. NOAA serves as the ISCCP long-term archive and processing center for the polar orbiting data. In these capacities, NOAA is currently involved in an effort to make ISCCP raw radiance data sets more accessible.

NOAA is pleased to join in this important dialog with EUMETSAT and other CGMS members as we enter a new era in the use of operational environmental satellite data for climate monitoring and diagnosis. Although operational environmental satellite data have been used in climate research for several decades, the growing importance of long-term monitoring is leading to a transition of climate monitoring and diagnosis from research to operations. We are pleased that CGMS is responding to this change and NOAA looks forward to conitinuing discussions in the area of climate applications.

2. PRINCIPLES OF SCIENTIFIC DATA STEWARDSHIP

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 responsibility for long-term archiving of the United States environmental data and has recently integrated several data management functions into a concept called Scientific Data Stewardship. Scientific Data Stewardship a new paradigm in data management consisting of an integrated suite of functions to preserve and exploit the full scientific value of NOAA's, and the world's, environmental data These functions include careful monitoring of observing system performance for long-term applications, the generation of authoritative long-term climate records from multiple observing platforms, and the proper archival of and timely access to data and metadata.

NOAA has developed a conceptual framework to implement the functions of scientific data stewardship. This framework has five objectives:

1) develop real-time monitoring of all satellite observing systems for climate applications,

2) process large volumes of satellite data extending up to decades in length to account for systematic errors and to eliminate artifacts in the raw data (referred to as fundamental climate data records, FCDRs),

3) generate retrieved geophysical parameters from the FCDRs (referred to as thematic climate data records TCDRs) including combining observations from all sources,

4) conduct monitoring and research by analyzing data sets to uncover climate trends and to provide evaluation and feedback for steps 2) and 3), and

5) provide archives of metadata, FCDRs, and TCDRs, and facilitate distribution of these data to the user community.

The term 'climate data record' and related terms, such as climate data set, have been used for some time, but the climate community has yet to settle on a concensus definition. A recent United States National Academy of Sciences report recommends using the following definition: a climate data record (CDR) is a time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change.

Furthermore, this report recommends a hierarchy of CDRs. Fundamental CDRs (FCDRs) are sensor data (e.g., calibrated radiances, brightness temperatures, radar backscatter, etc.) that have been improved and quality-controlled over time, together with the ancillary data used to calibrate them. Thematic CDRs (TCDRs) are geophysical variables derived from the FCDRs, specific to various disciplines, and often generated by blending satellite observations, in-situ data, and model output.

Each phase of this end-to-end system of Scientific Data Stewardship 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 internationally. In the following sub-sections, we assess current activities by NOAA and the international climate community and respond to recommendations by EUMETSAT made in CGMS-XXXI EUM-WP-22 within the framework of Scientific Data Stewardship .

1*) Develop real-time monitoring of all satellite observing systems for climate applications*

CGMS has sponsored an effort to operationally intercalibrate the geostationary imagers for several years for real-time weather applications. Additionally, the large operational numerical weather prediction centers also monitor the satellite calibration and this information is also valuable in monitoring instrument performance. These monitoring activities are important in the early identification of instrument calibration issues and make reprocessing easier. Such efforts are necessary, but not sufficient for long-term climate monitoring. The ISCCP has sponsored an intercalibration for long-term climate research applications, but this calibration requires collection of month to years of data and several updates of the calibration coefficients over time.

Recommendation: CGMS members should explore methods for collecting, archiving and making accessable a suite of intercalibration analysis including those made in real-time, for climate monitoring, and from NWP centers such as NCEP and ECMWF.

Experience has shown that, over time, satellite instruments may be subject to calibration drifts, operating at temperatures outside of the pre-launch specifications, and other anomalous conditions. Because of this, it can be expected that the calibration will need to be updated at some future date. This means that the calibration process should be reversible at some fundamental level for the satellite data sets in an easy to use format.

Recommendation: Satellite operators should provide the raw data (i.e., FCDRs) in an easy to use format that provides a reversible application of the calibration information between fundamental instrument observations and calibrated output with all information necessary for the transformations included as metadata.

2) Process large volumes of satellite data extending up to decades in length to account for systematic errors and to eliminate artifacts in the raw data (referred to as fundamental climate data records, FCDRs.

EUMETSAT mentions its Climate Data Set (CDS) and the ISCCP and GCPC data sets. These are all important efforts and analysis of these data sets have resulted in many advances in our understanding and modeling of the climate system. In particular, ISCCP has paid particular attention to the generation of FCDRs. NOAA is currently working to facilitate the use of the ISCCP FCDR level B1 geostationary data (raw data from all geostationary satellites subsampled at 3 hours and 10 km; this is the only source of GMS geostationary data from 1983-1987) and level B2 polar-orbiting data. Subsampled data sets help reduce volume and make it possible for many more users to analyze the data. Numerous other satellite data sets could be subsampled and combined with the operational geostationary and polar-orbiting satellite data, and in situ data, to create a 'quilt' of integrated data sets for climate studies.

Recommendation: CGMS members should review current combined, subsampled data sets and user requirements and consider establishing integrated FCDRs for the operational environmental satellites and relevant in situ data.

3) Generate retrieved geophysical parameters from the FCDRs (referred to as thematic climate data records TCDRs) including combining observations from all sources.

The production of TCDRs is more problematic than the production of FCDRs bacause of the varied uses of climate data, the complexities of data generation, and the difficulties of sustaining the program over extended periods of time. NOAA encourages the development of novel TCDRs by interested groups. The process for deciding what specific TCDRs should be generated by operational agencies is complex and varies from agency to agency. NOAA is currently involved in a process with a United States National Academies panel to solicit their input and advice on how to best approach the generation of TCDRs. This panel has made several detailed recommendations within the broad areas of organization, generation and stewardship, and sustainability of TCDRs. NOAA will be assessing these recommendations and will be preparing an implementation plan for the generation of TCDRs later in 2004. NOAA anticipates that the organizational portion of the implementation plan will include a framework for mechanisms for scientific oversight and advice, encouraging feedback from user communities, and allowing opportunities to re-direct the program

based upon advice and feedback. CGMS members will be one of many user groups consulted in this process.

Recommendation: CGMS members are encouraged to develop novel climate data sets (referred to as Thematic Climate Data Records, TCDRs), but international coordination efforts should primarily concentrate on the generation of subsampled, combined FCDR data sets in collaboration with both the operational and research climate communities.

4) Conduct monitoring and research by analyzing data sets to uncover climate trends and to provide evaluation and feedback for steps 2) and 3).

Monitoring of both FCDRs and TCDRs is an important function but, similar to the generation of TCDRs, varies greatly from agency to agency. NOAA through the reanalysis efforts at the Environmental Modeling Center is engaged in long term trend analyses of the Global Observing System including the satellite component; this reanalysis is conducted periodically in an attempt to reprocess archived data with the current algorithms.

5) Provide archives of metadata, FCDRs, and TCDRs, and facilitate distribution of these data to the user community.

NOAA believes that CGMS can and should play a very important role in the coordination of operational satelite data and metadata. NOAA endorses the initial recommendations of EUMETSAT provided in paper EUM-WP-22 and additional recommendations added at CGMS XXX WG II and AOPC IX. Archival and metadata issues, however, extend far beyond just the members of CGMS. It is critical that CGMS actively participate in and adopt the metadata standards being developed by the International Organization for Standardization (ISO).

Understanding complex environmental problems demands data from many observing systems. In order to be useful, these data need to be documented in ways that many users can understand. As part of the U.S. Federal Government, NOAA is required to document all new geospatial data collected or produced using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (see http://www.fgdc.gov).

Three metadata standards need to be implemented for NESDIS and NOAA: the FGDC Content Standard for Digital Geospatial Metadata (Version 2), the FGDC Extensions for Remote Sensing Metadata (FGDC-STD-012-2002), and ISO Standard 19115. These three standards are very similar and will converge in the short term (likely CY 2004/05). Crosswalks that allow migration of content between the standards are being actively developed with the support of the FGDC.

The standard includes seven sections:

Section	Description
1. Identification	Basic information about the data (mostly data discovery): citation, description, time period of content, status, spatial domain, keywords (theme, place), access constraints, use constraints, point of contact, browse graphic, data set credit, security, native dataset environment, cross reference
2. Quality	Attribute accuracy, logical consistency, completeness, positional accuracy, lineage, cloud cover

3. Spatial Data Organization	Indirect spatial references, direct spatial references (raster object information)
4. Spatial Reference	Horizontal coordinate system (projection), vertical coordinate system
5. Entity and Attribute	Detailed description (entity type, i.e. grid cell and attributes / parameters), Overview description
6. Distribution	Distributor, resource description (URL), distribution liability, standard order process (formats), custom order process, technical prerequisites, available time period
7. Metadata	Metadata date, review date, future review date, contact, metadata standard, metadata standard version, time convention, metadata access and use constraints, security, extensions

The Remote Sensing Extensions (RSE) define content standards for metadata, not defined in the FGDC Metadata Content Standard, that are needed to describe data obtained from remote sensing activities. These extensions are clearly critical to describing many of the datasets created and archived by NESDIS. They include metadata describing the sensor, the platform, the method and process of deriving geospatial information from the raw telemetry, and the information needed to determine the geographical location of the remotely sensed data. In addition, metadata to support aggregation, both the components of an aggregate data set and the larger collection of which a data item may be a member, are supported.

The ISO 19115 standard (scheduled to be finalized during 2004) includes many of the same sections as the FGDC and some important new information. For example, the ISO standard includes a much more useful way to describe a URL than the FGDC standard. It includes a name and a description for the URL in addition to a category of the service provided by the URL (download, information, offline Access, order, or search). This allows data providers to create meaningful Web pages directly from the metadata and to link to more detailed search mechanisms from a general data discovery system. The ISO standard includes richer code sets for describing associations between datasets. The ISO also includes several mechanisms that support aggregation of metadata records.

Recommendation: CGMS members should adopt the metadata standard ISO 19115 with appropriate extensions when that standard is finalized

Adoption of ISO 19115 will not only ensure all metadata recommended in CGMS-XXXI EUM-WP-22 be available, but a much richer set of metadata will be available in a standard format that ensure ease of access and use across a wide range of scientific and technical disciplines.

4 CONCLUDING REMARKS

CGMS XXXII is invited to:

- note the work at NOAA on climate data from operational meteorological satellite
- make, if possible, some general recommendations on the production of climate data from satellites embracing all CGMS members
- note of the concurrence of GCOS-AOPC with the list of meta-data that must be available with satellite measured data
- amend the list of meta-data to include instrument and data processing information.