CGMSXXX
USA-WP-27
Agenda Item: II.6
Discussed in WG-2

Current and future plans for metadata for the NOAA Geostationary Operational Environmental Satellites (GOES)

Summary and Purpose of Document

This paper summarizes the current and planned metadata attached to archived GOES data. This is the USA response to Action 29.33

Action Requested: None

CURRENT AND FUTURE PLANS FOR METADATA FOR THE NOAA GEOSTATIONARY OPERATIONAL ENVIRONMENTAL SATELLITES (GOES)

John J. Bates [1], Chris Velden [2], and Timothy J. Schmit [3]

[1] NOAA/NESDIS/NCDC
Asheville, NC

[2] Cooperative Institute for Meteorological Satellite Studies
[3] NOAA/NESDIS/ORA/ASPT
Madison, Wisconsin, 53706 USA

1. Introduction

In its most generic definition, metadata refers to the ancillary information about a given data set. In the past, metadata saved as part of the GOES satellite data stream was only rudimentary, usually only enough metadata was saved to describe the data file. Additional information about data quality may have been saved by the instrument operator or ingest facility, but such data was largely inaccessible to the general user community.

With the increased interest in the use of satellite data for long-term climate studies more recently, the demands for additional metadata and tools to easily access and manipulate the metadata have greatly increased. NOAA has recognized the need for the production of additional metadata from past satellites, the need for making such metadata easily available to users, and the need for operational implementation of enhanced metadata for current and future satellites. The following sections describe NOAA's efforts in these areas.

2. Current metadata available from the GOES satellites

The metadata available currently for the GOES series satellites is limited. Current metadata includes: satellite id, Julian day, calendar day, time (GMT), satellite schedule (e.g., routine, rapid scan, special), sector coverage (e.g., NH=northern hemisphere, FD=full disc, CONUS=continental United States, etc.), and four corners of coverage area in latitude and longitude. Some additional information on data quality (e.g., number of dropped scan lines) are kept at the University of Wisconsin – Madison Space Science and Engineering Center, but are not generally available at this time.

3. The Comprehensive Large-Array Stewardship System (CLASS) Satellite Metadata Characterization project

The NOAA CLASS project is characterized as:

- Main entry portal for NOAA/NESDIS environmental data
- Archive, Access and distribution system primarily for NOAA's seven large array data sets (POES/DMSP, GOES, EOS, NPP, Metop, NEXRAD, & NPOESS)
- A reengineering effort rather than a totally new development effort
- A phased implementation based upon requirements and critical dates

As part of the CLASS program, NESDIS is copying all the GOES data originally archived at the University of Wisconsin – Madison SSEC to modern media and an automated mass storage system. The copying of the archive files is proceeding well and is scheduled to be completed by the middle of 2003. Parallel with this effort is a project to allow users access to these archive, as well as future, GOES data. This project, called the GOES active archive (GAA), will allow user access to all historic GOES data through a user interface similar to that provided by the current satellite active archive (SAA) for NOAA polar-orbiter data. The first full release of the GAA is also

scheduled for the middle of 2003. Access to the GAA will be provided by mirror sites at the National Climatic Data Center (NCDC) in Asheville NC and the Office of Satellite Data Processing and Distribution (OSDPD) in Suitland, MD.

Within the framework of the CLASS, the satellite metadata characterization project has been established to provide information critical to data discovery, usability, quality, interoperability and automatic processing. A strong metadata management system is critical to the success of the NESDIS satellite data systems. NESDIS Satellite Metadata is presently spread across a number of locations and systems and does not generally take advantage of national (Federal Geographic Data Committee, FGDC) and international (International Standards Organization, ISO) metadata standards. An important goal is to understand these systems and evaluate the advantages and disadvantages of coalescing them.

As a step toward that goal we have collected definitions for the fields in six existing satellite metadata collections. We have also created crosswalks that identify fields from different metadata collections that have similar meanings (semantics). We designed a database for holding the field definitions and describing the crosswalks and a web-based system for populating the definitions and creating the crosswalks. We are presently working with the NESDIS satellite metadata processing groups to check the definitions and crosswalks we have created. Once that is done, we will be using the crosswalks to coalesce information in all six collections into an FGDC compliant metadata collection that describes several hundred NESDIS metadata products.

4. Production of improved metadata within the NOAA Scientific Data Stewardship (SDS) Program

The NESDIS scientific stewardship program has five goals. These goals are:

Provide real-time monitoring of climate-scale biases in the global suite of satellite observing systems. Since subtle spatial and temporal biases can create serious problems in future use of the data, we must develop the tracking tools necessary for detection of biases in the climate record. These biases can then be minimized or eliminated through efficient communication and coordination of information related to network performance using both in situ and satellite observations.

Document Earth system variability and change on global, regional, and local scales. This will be accomplished by building and maintaining a high quality base of data and information and establishing the best possible historical perspective critical to effective analysis and prediction. The creation of long-term, consistent records requires a long-term commitment of resources to accomplish these tasks.

Provide the necessary algorithms to ensure that understanding of key climate processes can be derived from space-based systems and the combination of space-based and in situ systems. The best possible scientific understanding of critical climate and global change issues can only be reached when all opinions and ideas can be explored. Thus, an active program engaging the research community, partnerships with industry, and increased interactions with local and regional governments are envisioned.

Optimize data and information services in order to make research easier and more effective by ensuring those services are simple, straight forward, direct, and responsive. This will be achieved by establishing end-to-end accountability for establishing long-term, scientifically valid, and consistent records for global change studies. This will ensure that our data and information are available to the maximum amount of users

Enable and facilitate future research. Because action is required now, but climate and global change societal imperative questions may not come into focus for many years, we must invoke the concept of stewardship to justify this effort. This aspect of stewardship involves providing the

basic information technology, hardware, telecommunications, and software support to guarantee that the data can be safeguarded and communicated both within NOAA and to outside users for generations in the future.

We believe that achieving these goals will result in a long-term archive that is flexible and innovative, that appropriately focuses responsibility on NESDIS for preservation of optimal data character, that provides for open access to the data by the scientific community and the public, and that will rapidly track technological developments.

As part of the availability of all GOES historical data through the GAA, the NCDC will be processing through the archive to construct metadata summaries of statistical properties of each sector. The statistical summaries will provide, for each channel, the mean, standard deviation, skewness, kurtosis, maximum, minimum, total number of good observations, and total number of missing and bad observations.

5. An example of metadata needed to generate GOES sounder temperature / moisture / cloud retrievals

From necessary to desired for temperature / moisture retrievals:

Multi-band radiance information
Radiance bias (each band) -- possibly built from co-located radiosonde data
Initial Temperature/moisture (and ozone) profiles for same time as radiance information
Surface analysis (sea-level-temperature, moisture, 1000 hPa height)
SST estimate (e.g, from AVHRR)
Regression coefficients for heating/cooling model (for cloud-clearing)
Calibration change information (patch temperature, visible degradation, etc)

From necessary to desired for cloud retrievals:

Multi-band radiance information
Calibration change information
Radiance bias (each band)
Surface analysis (temperature, moisture, 1000 hPa height)
SST estimate (e.g., from AVHRR)
Initial Temperature/moisture profiles
Regression coefficients for heating/cooling model

Code needed:

RTM and coefficients
Retrieval codes
Auxiliary files (inverse of co-variance matrix, etc.) are needed for
temperature/moisture retrievals
Cloud mask