HDF5 FOR NPOESS SENSOR AND ENVIRONMENTAL DATA RECORDS

Richard Ullman[1], Ronald Andrews[2]

[1]NASA Goddard Space Flight Center, Code 586, Greenbelt, MD; [2]Northrop Grumman Aerospace Systems Corp., One Space Park R10/1342, Redondo Beach, CA

1. INTRODUCTION

The National Polar-orbiting Operational Environmental Satellite System (NPOESS) is the next generation of low earth orbiting environmental satellites. The NPOESS and the NPOESS Preparatory Project (NPP) satellite are sun-synchronous polar orbiters with a period approximately once every 100 minutes. Together with the Interface Data Processing Segment (IDPS) the system will provide global monitoring environmental conditions, collecting, disseminating and processing data about the Earth's weather, atmosphere, oceans, land, and near-space environment with precision and detail never before achieved by operational weather satellites. This volume of data will allow scientists and forecasters to monitor and predict weather patterns with greater speed and accuracy.

NPP/NPOESS Data products are delivered as HDF5 files. HDF5 is a general-purpose file format and library designed and developed by the National Center for Supercomputing Applications (NCSA). HDF5 was developed to provide flexible, portable and efficient storage and retrieval of scientific datasets. NPOESS uses the HDF5 structure to implement a specific data model for the NPP/NPOESS data products without use of any required extensions to the native library. Some advantages of using HDF include efficient storage and I/O, including parallel I/O and the fact that it is free, open source software, available for use on multiple platforms [1].

2. NPOESS DATA AND METADATA ENCODING

Utilizing mature technology standards from both HDF5 and XML, NPP/NPOESS data products provide platform mobility and accessibility by a diverse set of users. In keeping consistency in mind for general application and framework development by the varied community of users, the NPOESS has strived to provide a common, consistent structure to the NPP/NPOESS data products HDF5 organization. Additionally the structure and individual NPP/NPOESS data products are fully described in publicly available documentation as well as machine-readable XML files, referred to as Product Profiles. The NPP/NPOESS products contain metadata (including real-time quality information), structured dataset arrays, facilitate aggregations of granules, and geolocation data [2].

Metadata is facilitated through detailed documentation ranging from Algorithm Documents and Data Format Control Books, describing the more static, consistent attributes of the data, to the attributes provided in the HDF5 files, describing the dynamic product instance. In addition to the product documentation and dataset attributes (field metadata), NPP/NPOESS Data products also provide quality flags (element metadata) via bit-fields co-aligned with the datasets.

3. PRODUCT PROFILES

NPP/NPOESS Product Profiles are encoded in XML and are distributed with the NPOESS documentation set. Each instance of a product type has a separate profile that is linked to the data granule through a metadata reference within the HDF5 product files. The XML files provide detailed information such as units of measure, dimension names, and legend entries that do not change from instance to instance. EDR and SDR profiles follow an XML schema; facilitating machine parsing for extracting desired elements. NPOESS provides a style sheet for rendering the Product Profiles in a web browser.

Each NPP/NPOESS Data Product is processed and packaged in HDF5 as structured arrays. Fields are stored as HDF5 datasets using HDF5 native data types and explicit array dimensions. Datasets within a single product that contain common dimensions are related by congruency. The dimensions and other static attributes of the arrays are provided via the XML Product Profiles.

4. GRANULES AND AGGREGATES

The granule is the atomic unit, or smallest sub-set of data, for the NPP/NPOESS Data products. Granules are time-based durations of the data produced from the senor output. Taking advantage of the hierarchical nature of HDF5, NPP/NPOESS Data Product granules can be aggregated together in a single HDF5 file without modification to the base structure of the HDF5 implementation designed for NPP/NPOESS Data products. Aggregation of granules is in the "along-track" dimension, by simply extending along that dimension. A pointer structure provides the ability to access the data in an aggregation by way of individual granules or as the whole dataset.

5. GEOLOCATION

Geolocation data for NPP/NPOESS Data products is constructed using the same paradigm and conventions as the science/sensor data they are associated with. This data, like the quality flags and other datasets in an NPP/NPOESS Data Product, have a congruence relationship with the same dimensions as the datasets to which they apply. Based on the HDF5 design employed for NPP/NPOESS Data products, geolocation data can be stored in the same HDF5 file as the science/sensor data in either a separate HDF5 group or in a separate file.

In this presentation, we will discuss the implementation of the various features described above and provide practical examples and references

6. REFERENCES

[1] The HDF Group. *Why Use HDF?*. Retrieved December 18, 2009, from http://hdfgroup.org/why_hdf/. September 23 2009.

[2] Purcell, P. *NPOESS Common Data Format Control Book – External Volume I – Overview.* (Document No. D34862-01). Northrop Grumman Aerospace Technology Corp. Redondo Beach, CA. 2009.