

HDF5 for NPOESS

- Hierarchical Data Format 5 (HDF5) is the format for delivery of processed products from the National Polar-orbiting Operational Environmental Satellite System (NPOESS) and for the NPOESS Preparatory Project (NPP).
- HDF5 is a general purpose library and file format for storing scientific data. Two primary objects:
 - Dataset, a multidimensional array of data elements
 - Group, a structure for organizing objects
- Efficient storage and I/O, including parallel I/O.
- Free, open source software, multiple platforms.
- Data stored in HDF5 is used in many fields from computational fluid dynamics to film making.
- Data can be stored in HDF5 in an endless variety of ways, so it is important to standardize how NPOESS product data is organized in HDF5.

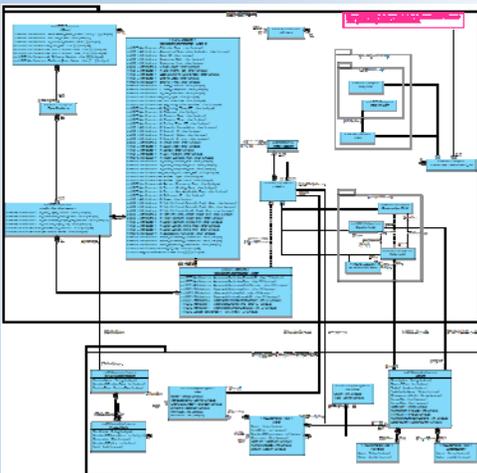
Format Strengths

- Straight HDF5.
 - No need for additional libraries.
- Consistent HDF5 group structure
 - Organization for each product is the same as all others.
 - Data "payload" is always in a product group within All_Data group.
- Allows for flexible temporal aggregation
 - Granules are appended by extending dataset dimension.

Format Challenges

- Geolocation appears in a separate product group and may be in separate HDF5 file.
- Field metadata, used to interpret data (similar to netCDF CF) are in separate product profile file.
- Quality flags must be parsed before they can be interpreted.
- Information needed for un-scaling scaled integers is not obvious.
- HDF5 indirect reference link API, used to link metadata to the data in NPOESS' use is complex and not supported by all analysis COTS implementations.

Information Model UML Diagram

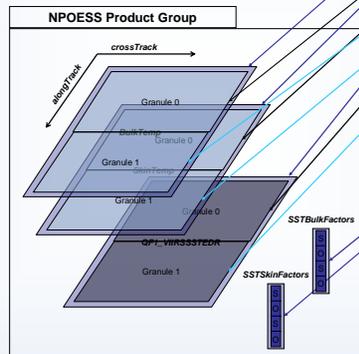


An Example Product Group

- In this example product group:
 - Five datasets constitute the product.
 - There are two common dimensions.
 - There are three congruent datasets.
 - Two datasets contain scale and offset values.
 - One dataset contains quality flags by element.
 - There are two granules in this aggregation.
 - Dimension "alongTrack" crosses the "granule boundary."

| Example extracted from VIIRS Sea Surface Temperature EDR | | | | |
|--|------------------------------|-------------------------|---------------|-------------------|
| Field Name | Description | Data Type | Dimensions | Units |
| BulkTemp | Sea Surface Bulk Temperature | 16-bit unsigned integer | [N*768, 3200] | Kelvin / Unitless |
| SkinTemp | Sea Surface Skin Temperature | 16-bit unsigned integer | [N*768, 3200] | Kelvin / Unitless |
| QF1_VIIRSSSTEDR | Land/Water Background | 1-bit | [N*768, 3200] | Unitless |
| | SST Skin Quality | 2-bit | | Unitless |
| | SST Bulk Quality | 2-bit | | Unitless |
| | Aerosol Correction | 3-bit | | Unitless |
| SSTBulkFactors | Bulk SST Scale | 32-bit float | [N*2] | |
| | Bulk SST Offset | 32-bit float | | Kelvin |
| SSTSkinFactors | Skin SST Scale | 32-bit float | [N*2] | Unitless |
| | Skin SST Offset | 32-bit float | | Kelvin |

N in Dimension is number of granules



| Geolocation Content Common Elements for VIIRS Geolocation Products | | | | |
|--|----------------------|---------------------|---------------|-----------------------|
| Field Name | Comments | Dimensions | Units | Data Type |
| StartTime | since epoch 1/1/1958 | [per scan or swath] | microsecond s | 64-bit signed integer |
| MidTime | since epoch 1/1/1958 | [per scan or swath] | microsecond s | 64-bit signed integer |
| SCPosition | ECR coordinates | [per scan or swath] | meters | 32-bit float |
| SCVelocity | ECR coordinates | [per scan or swath] | meters/second | 32-bit float |
| Latitude | | [per cell] | degrees | 32-bit float |
| Longitude | | [per cell] | degrees | 32-bit float |
| SolarZenithAngle | | [per cell] | degrees | 32-bit float |
| SolarAzimuthAngle | | [per cell] | degrees | 32-bit float |
| SensorZenithAngle | | [per cell] | degrees | 32-bit float |
| SensorAzimuthAngle | | [per cell] | degrees | 32-bit float |
| Height | geoid or terrain | [per cell] | meters | 32-bit float |
| SatelliteRange | | [per cell] | meters | 32-bit float |

Dimensions

- Dimensions are defined for each field.
- Fields are related by congruency and common dimensions.
- Common dimensions are given the same name.
- One dimension crosses the granule boundary. When multiple granules are "aggregated" the "granule boundary" dimension is extended.
- Dimension names and attributes are provided in the product profile.

Granules and Aggregates

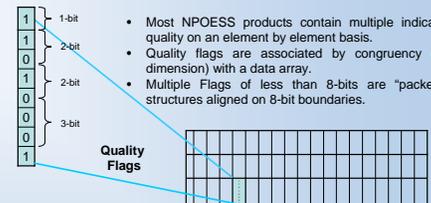
- Aggregate (all fields)
 - complete datasets in All_Data Group
- Granule 0 (all fields)
 - "region" in each dataset in All_Data Group
- Granule 1 (all fields)
 - "region" in each dataset in All_Data Group

Scaled Integer Storage

- For storage efficiency floating point data values may be stored as scaled integers.
- To re-generate the data value, the dataset element must be multiplied by a supplied scale factor and an integer offset added.
- The scale factor and offset are provided, one pair for each granule as a separate dataset.
- The scale and offset value is the same for all granules produced with a given version of an algorithm - not dynamic scaling.
- The fact that a dataset is a scaled value and the association between the data dataset and the scale factor dataset is contained in the product profile.

Quality Flags by Element

- Most NPOESS products contain multiple indicators of quality on an element by element basis.
- Quality flags are associated by congruency (shared dimension) with a data array.
- Multiple Flags of less than 8-bits are "packed" into structures aligned on 8-bit boundaries.



Geolocation

- Geolocation products are constructed using the same conventions as SDRs and EDRs.
- Geolocation datasets have a congruency relationship with the same dimensions as the datasets to which they apply.
- The association between a data product with its geolocation product is made on one of two ways:
 - The geolocation product may be packaged as a separate product group within the same HDF5 file.
 - The name of a separate geolocation product file may be stored in the N_GEO_Ref attribute on the root HDF group.
 - Choice of "as a product group" or "as a separate file" is made upon order from the NPOESS.