

RFC: Writing User Block for NPP files with nagg

Larry Knox
Elena Pourmal

This document describes a new feature to write NPP metadata elements in a user block in nagg output files, according to the specification in the JPSS Common Data Format Control Book – External – Volume V, Section 3.2.

Table of Contents

1	Introduction.....	2
2	User block requirements	2
3	Implementation considerations.....	3
3.1	Top level elements.....	3
3.2	Data_Product elements.....	3
3.3	Nagg structures for user block element values	4
4	Unaddressed potential issue	4
5	Testing for user block	4
6	WBS with estimates	Error! Bookmark not defined.
7	Summary and Recommendations	5
8	Revision History	5
9	Appendix 1 - HDF5 SDR, TDR, EDR, ARP, GEO, and IP HDF5 XML User Block Schema (JPSS CDFCB – External – Volume V, Section3.2.1).....	5
10	Appendix 2 - User block example with packaged products	6
10.1	Section 1 (Yellow)	7
10.2	Section 2 (Blue, Olive)	8

1 Introduction

Nagg (http://www.hdfgroup.org/projects/npoess/nagg_index.html) is a tool for aggregating JPSS data products from existing files into new files with a different number of granules per file or different combinations of compatible products. JPSS data files have a user block that contains XML elements which are a subset of the metadata elements in the file, but can be accessed without HDF5 tools or libraries. These elements are attributes of the root group, the /Data_Products/<Collection Short Name> groups, or the /Data_Products/<Collection Short Name>/<Collection Short Name>_Aggr datasets in the file. The user block for the new file should contain these elements with the appropriate values for the products and aggregations in the new file. This RFC enumerates the currently required elements of this user block and the locations of the corresponding attributes for obtaining the element values.

While the user block element names match the names of the corresponding attributes, the location of those attributes is not given in the user block, and the values of the aggregation's attributes are determined as the file is written. Although the current JPSS specification for the user block appears to be static, nagg should read the user blocks of the input files to check for unexpected new elements or user additions and issue a warning if any are found, as they will not be written in the output file.

2 User block requirements

The XML schema for JPSS data product files of type SDR, TDR, EDR, ARP, GEO, and IP is specified in the JPSS Common Data Format Control Book – External Volume V – Metadata, section 3.2.1, and is included in Appendix 1.

The user block will contain exactly one of each of the first 4 elements and 1 or more of the Data_Product elements to match the number of JPSS data products included in the file. While nagg does not currently support aggregation of RDR data products, note for future reference that the user block for RDRs differs only by the absence of the N_GEO_Ref element.

Each Data_Product element will contain exactly one of the following elements:

- N_Collection_Short_Name
- Instrument_Short_Name
- N_Dataset_Type_Tag
- N_Processing_Domain
- AggregateBeginningDate
- AggregateBeginningOrbitNumber
- AggregateBeginningTime
- AggregateEndingDate
- AggregateEndingOrbitNumber
- AggregateEndingTime
- AggregateBeginningGranuleID
- AggregateEndingGranuleID

An example of a user block for a file with 2 products is included in Appendix 2.

3 Implementation considerations

Nagg copies the values of the attributes corresponding to the user block elements from the input file to the output file in the course of processing the files. It should be most efficient to read and save the values for the user block at that time or to replace the copy by reading and writing them to the attribute and the user block string at the time they are copied from the input granule. The required size of the user block can be calculated before creating the file according to the element names and types and the number of products to be packaged in the output file. Element values can be saved for writing after the product granules are processed for an output file.

A list of expected elements will be useful for checking the input files' user blocks. If more flexibility is required in the future, this list could be put in a configuration file to be read at runtime. Entries would need an HDF5 path for the corresponding attribute. This is not part of the work currently proposed.

3.1 Top level elements

`Mission_Name` and `Platform_Short_Name` are root group attributes as well as user block elements. Their values should be the same for all input and output files.

The `N_GEO_Ref` element for the output file user block will have the same value as the `/N_GEO_Ref` attribute. If the geolocation is packaged, the `/N_GEO_Ref` attribute will not be created. The user block element will be present, but empty (`"<N_GEO_Ref></ N_GEO_Ref>"`).

The value of the `Number_of_Data_Products` element will be determined by the number of data products packaged in each output file (including the geolocation product).

3.2 Data_Product elements

The first four elements correspond to a subset of the attributes of the `/Data_Products/<Collection Short Name>` group which use the same names:

```
N_Collection_Short_Name
Instrument_Short_Name
N_Dataset_Type_Tag
N_Processing_Domain
```

These values are copied by nagg **from the first input file** to the output file.

The remaining eight elements correspond to a subset of attributes of the `/Data_Products/<Collection Short Name>/<Collection Short Name>_Aggr` dataset:

```
AggregateBeginningDate
AggregateBeginningOrbitNumber
AggregateBeginningTime
AggregateEndingDate
AggregateEndingOrbitNumber
AggregateEndingTime
AggregateBeginningGranuleID
AggregateEndingGranuleID
```

The values for these attributes are copied by nagg from the first and last granules in the aggregation. They can also be read and written to the user block.

3.3 Nagg structures for user block element values

The nagg structures to be added will be similar to these. Character arrays can be replaced by char *s where information is obtained from persistent granules.

```
struct user_block_t {
    char mission_name[];
    char platform_name[];
    char *geofile_name;
    unsigned int number_data_products;
    struct data_product_info_t *data_products[];
}

struct data_product_info_t {
    char product_name [];
    char instrument_name[];
    char dataset_type_tag[];
    char processing_domain[];
    char beginning_date[];
    uint64_t beginning_orbit;
    char beginning_time[];
    char ending_date[];
    uint64_t ending_orbit;
    char ending_time[];
    char beginning_granule_id[];
    char ending_granule_id[];
}
```

4 Unaddressed issue

The value of /N_Processing_Domain may not be the same for all input files to be aggregated. The value of the root group attribute for the output file is copied from the root group attribute in the input file for the first granule to be aggregated. If later granules in the same aggregation are from a file with a different value for /N_Processing_Domain, that value will be ignored for the current aggregation. There is no provision for this situation, which applies to several other attributes in the CDFCB specification.

5 Testing for user block

A test program should be written to read the xml tags in the user block and compare the values to those found by reading the corresponding attributes of the root group, the product groups, and the <product>_aggr datasets. A utility to output the user block with character returns and indentations as displayed in the appendix would also be useful.

6 Summary and Recommendations

Nagg currently produces files without information expected in the user block. The values can be determined when nagg is creating new files, and for most nagg operations the majority of the values in the user block must be determined when creating the output file, as they will be different from the values in the input files. The user block should be added with the implementation suggested by the information in section 3.

7 Revision History

September 20, 2013: Version 1 available for internal review.

Version 2 committed to subversion repository

8 Appendix 1 - HDF5 SDR, TDR, EDR, ARP, GEO, and IP HDF5 XML User Block Schema (JPSS CDFCB – External – Volume V, Section 3.2.1)

```
<?xml version="1.0"?>
<xsd:schema xmlns:xsd=http://www.w3.org/2001/XMLSchema targetNamespace=http://
"NPOESS SDR, TDR, EDR, ARP, and IP XML User Block" xmlns="http://NPOESS
SDR, TDR, EDR, ARP, and IP XML User Block">
<!-- == NPP/NPOESS HDF5 XML User Block XML Schema == -->
<!-- == NPP/NPOESS HDF5 User Block - Primary Element== -->
<xsd:element name="HDF_UserBlock" type="HDF_UserBlockType" />
<!-- == Annotation for this schema == -->
<xsd:annotation>
<xsd:documentation xml:lang="en">
NPP/NPOESS HDF5 User Block XML Schema
</xsd:documentation>
</xsd:annotation>
<!-- == NPP/NPOESS HDF5 User Block == -->
<xsd:complexType name="HDF_UserBlockType">
<xsd:sequence>
<xsd:element name="Mission_Name" type="xsd:string" minOccurs="1"
maxOccurs="1" />
<xsd:element name="Platform_Short_Name" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="N_GEO_Ref" type="xsd:string"
minOccurs="0" maxOccurs="1" />
<xsd:element name="Number_Of_Data_Products" type="xsd:integer"
minOccurs="1" maxOccurs="1" />
<xsd:element name="Data_Product" type="Data_ProductType"
minOccurs="1" maxOccurs="unbounded" />
</xsd:sequence>
</xsd:complexType>
<!-- == Data Product Type == -->
<xsd:complexType name="Data_ProductType">
<xsd:sequence>
<xsd:element name="N_Collection_Short_Name" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="Instrument_Short_Name" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="N_Dataset_Type_Tag" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="N_Processing_Domain" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="AggregateBeginningDate" type="xsd:string" minOccurs="1"
```

```

maxOccurs="1" />
<xsd:element name="AggregateBeginningOrbitNumber"
type="xsd:integer" minOccurs="1" maxOccurs="1" />
<xsd:element name="AggregateBeginningTime" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="AggregateEndingDate" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="AggregateEndingOrbitNumber" type="xsd:integer"
minOccurs="1" maxOccurs="1" />
<xsd:element name="AggregateEndingTime" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="AggregateBeginningGranuleID" type="xsd:string"
minOccurs="1" maxOccurs="1" />
<xsd:element name="AggregateEndingGranuleID" type="xsd:string"
minOccurs="1" maxOccurs="1" />
</xsd:sequence>
</xsd:complexType>
</xsd:schema>

```

9 Appendix 2 - User block example with packaged products

Here is a user block example for a packaged file with 4 GMODE granules and 4 SVM01 granules:

```

<HDF_UserBlock><Mission_Name>NPP</Mission_Name><Platform_Short_Name>NPP</Platform_Short_Name><N_GEO_Ref></N_GEO_Ref><Number_Of_Data_Products>2</Number_Of_Data_Products><Data_Product><N_Collection_Short_Name>VIIRS-M1-SDR</N_Collection_Short_Name><Instrument_Short_Name>VIIRS</Instrument_Short_Name><N_Dataset_Type_Tag>SDR</N_Dataset_Type_Tag><N_Processing_Domain>ops</N_Processing_Domain><AggregateBeginningDate>20120813</AggregateBeginningDate><AggregateBeginningOrbitNumber>4115</AggregateBeginningOrbitNumber><AggregateBeginningTime>094839.337354Z</AggregateBeginningTime><AggregateEndingDate>20120813</AggregateEndingDate><AggregateEndingOrbitNumber>4115</AggregateEndingOrbitNumber><AggregateEndingTime>095419.727577Z</AggregateEndingTime><AggregateBeginningGranuleID>NPP000255233200</AggregateBeginningGranuleID><AggregateEndingGranuleID>NPP000255235761</AggregateEndingGranuleID></Data_Product><Data_Product><N_Collection_Short_Name>VIIRS-MOD-GEO</N_Collection_Short_Name><Instrument_Short_Name>VIIRS</Instrument_Short_Name><N_Dataset_Type_Tag>GEO</N_Dataset_Type_Tag><N_Processing_Domain>ops</N_Processing_Domain><AggregateBeginningDate>20120813</AggregateBeginningDate><AggregateBeginningOrbitNumber>4115</AggregateBeginningOrbitNumber><AggregateBeginningTime>094839.337354Z</AggregateBeginningTime><AggregateEndingDate>20120813</AggregateEndingDate><AggregateEndingOrbitNumber>4115</AggregateEndingOrbitNumber><AggregateEndingTime>095419.727577Z</AggregateEndingTime><AggregateBeginningGranuleID>NPP000255233200</AggregateBeginningGranuleID><AggregateEndingGranuleID>NPP000255235761</AggregateEndingGranuleID></Data_Product></HDF_UserBlock>

```

The same string with carriage returns between fields, indented for nesting is repeated below, color coded **yellow** for top level elements that have no sub-elements, **blue** for data product 1 elements, and **olive** for data product 2 elements.

```
<HDF_UserBlock>
```

<!--Root Group Attributes -->

<Mission_Name>NPP</Mission_Name>

<Platform_Short_Name>NPP</Platform_Short_Name>

<N_GEO_Ref></N_GEO_Ref>

<Number_Of_Data_Products>2</Number_Of_Data_Products>

No matching

attribute

<Data_Product>

<N_Collection_Short_Name>VIIRS-M1-SDR</N_Collection_Short_Name>

Product Group

attributes

<Instrument_Short_Name>VIIRS</Instrument_Short_Name>

<N_Dataset_Type_Tag>SDR</N_Dataset_Type_Tag>

<N_Processing_Domain>ops</N_Processing_Domain>

<AggregateBeginningDate>20120813</AggregateBeginningDate>

Aggregate Dataset

Attributes

<AggregateBeginningOrbitNumber>4115</AggregateBeginningOrbitNumber>

<AggregateBeginningTime>094839.337354Z</AggregateBeginningTime>

<AggregateEndingDate>20120813</AggregateEndingDate>

<AggregateEndingOrbitNumber>4115</AggregateEndingOrbitNumber>

<AggregateEndingTime>095419.727577Z</AggregateEndingTime>

<AggregateBeginningGranuleID>NPP000255233200</AggregateBeginningGranuleID>

<AggregateEndingGranuleID>NPP000255235761</AggregateEndingGranuleID>

</Data_Product>

<Data_Product>

<N_Collection_Short_Name>VIIRS-MOD-GEO</N_Collection_Short_Name>

<Instrument_Short_Name>VIIRS</Instrument_Short_Name>

<N_Dataset_Type_Tag>GEO</N_Dataset_Type_Tag>

<N_Processing_Domain>ops</N_Processing_Domain>

<AggregateBeginningDate>20120813</AggregateBeginningDate>

<AggregateBeginningOrbitNumber>4115</AggregateBeginningOrbitNumber>

<AggregateBeginningTime>094839.337354Z</AggregateBeginningTime>

<AggregateEndingDate>20120813</AggregateEndingDate>

<AggregateEndingOrbitNumber>4115</AggregateEndingOrbitNumber>

<AggregateEndingTime>095419.727577Z</AggregateEndingTime>

<AggregateBeginningGranuleID>NPP000255233200</AggregateBeginningGranuleID>

<AggregateEndingGranuleID>NPP000255235761</AggregateEndingGranuleID>

</Data_Product>

</HDF_UserBlock>

9.1 Section 1 (Yellow)

4 fields: Mission_Name

Platform_Short_Name

N_GEO_Ref

Number of Data_Products

The first 2 match attributes of the root group which are always present. The third matches a root group attribute that is present only for unpackaged files, but will be an empty field in the user block for files with packaged geolocation. The fourth is simply the number of data products packaged in the file.

9.2 Section 2 (Blue, Olive)

Section 2, Data_Product, is repeated once for each product in the file. It has 4 fields matching attributes of the Data_Products child group named with the value of the N_Collection_Short_Name field and 8 fields beginning with "Aggregate" matching attributes of the same group's Aggregate dataset.