

NAGG - NPP Aggregation and Packaging Tool Proposal for Prototype Implementation

The HDF Group

The HDF Group was tasked by JPSS to create an NPP granule aggregation and packaging tool. Due to the complexity of the task the development will be done in stages to assure that the HDF developers understand the requirements and implement the tool according to the specifications found in control books.

This document outlines the initial development steps and describes the prototype that will be delivered by December 21, 2011. It also specifies the future development steps leading to the full implementation.

1. Goal of the Prototype

The prototype we propose to implement will achieve two goals and will build a foundation for the full implementation.

The first goal is to discover information stored in the specified NPP files required for:

1. Creating an ordered list of the granules for all products found.
2. Creating new NPP files to store aggregated or de-aggregated granules.

The second goal will be to write a specified number of ordered granules of the same type to the new NPP files.

2. What will the Prototype Do?

The prototype will aggregate or de-aggregate one type of granule found in the specified NPP files according to the aggregation parameter provided by a user. It will also aggregate or de-aggregate the corresponding geolocation product. The tool will create new NPP files that will have file names and HDF5 directory structure according to the specifications found in the control books. The prototype will also create fill granules and write aligned files (buckets).

3. What will the Prototype not Do?

The prototype will not:

- Write a user block to a new NPP file
- Aggregate or de-aggregate NPP files that contain several products (in other words, packaging)
- Implement support for repair granules

- Implement the `-l` (lower case L) and `-S` tool flags
- Implement the `-A` flag

After the features described below have been implemented and tested, we propose to proceed with these additional capabilities.

4. Command Line Options

We propose to implement a set of the command line options that will allow us to achieve the goals specified in the “Goal of the Prototype” section above. The options are shown below:

```
nagg -n <number> -t <product_name> -d <directory> -O <origin> -D <domain>  
<filename>
```

where

- `<number>` is an integer greater than 0 indicating the number of granules in each aggregate file
- `<product_name>` is a name of one product (vs. the list of products)
- `<directory>` is name of the directory containing the output NPP files
- `<origin>` is the origin identifier, a four character string (required)
- `<domain>` is the domain identifier, a three character string (required)
- `<filename>` is a list of NPP files or a name of the directory with NPP files

5. Tools Architecture

The tools will consist of the following modules:

- Command line parser.
- Module to scan all input files and to produce a table with information about all granules including geolocation granules; see Section 5.1 for details.
- Module to sort a table with granules information; see Section 5.2 for details.
- Module to generate output file names (in other words, filenames according to control book specifications).
- Module to write granules from the ordered table; see Section 5.3 for details.

5.1. Gathering Information from the List of NPP Files

The tool will read the following information from the list of the specified NPP files and store it in an array of structures (in other words, a table) (TBD).

Parameter	Where to find the value	Comment
File name	In the list of the files or directory	
Product name or product type	As a name of a group stored under /Data_Products	Required for granule identification
Number of granules	The attribute "AggregateNumberGranules" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Aggr	This information is also found in the user block.
Granule start time (IET)(ULL)	The attribute "N_Beginning_Time_IET" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Gran_N	Also found in the user block as UTC date and time if a file contains only 1 granule
Granule end time (IET)(ULL)	The attribute "N_Ending_Time_IET" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Gran_N	Also found in the user block as UTC date and time if a file contains only 1 granule
Granule beginning date (CHAR)	The attribute "Beginning_Date" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Gran_N	Needed for output file name generation
Granule beginning time (CHAR)	The attribute "Beginning_Time" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Gran_N	Needed for output file name generation
Granule ending time (CHAR)	The attribute "Ending_Time" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Gran_N	Needed for output file name generation
Granule ID (CHAR)	The attribute "N_Granule_ID" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Gran_N	Required for granule identification; can be used for sorting granules and matching geolocation granules
Granule version (CHAR)	The attribute "N_Granule_Version" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Gran_N	Required for granule identification. When more than one granule with the same ID is found, the granule with the latest version will be used.
Name of the geolocation file (CHAR)	The attribute "N_GEO_Ref" stored in the root group.	This isn't present for the geolocation product.

Parameter	Where to find the value	Comment
Orbit number (ULL)	The attribute "N_Beginning_Orbit_Number" in the dataset /Data_Products/<CollectionShortName>/<CollectionShortName>_Gran_N	Needed for output file name generation

5.2. Sorting Information

The information stored in the array of structures will be sorted in ascending order by:

- Granule ID
- Product ID
- Granule version

5.3. Writing an Ordered List of Granules

In the earliest implementation of the prototype we assume that

- There are no gaps among the granules of the same type (in other words, there is no need for fill granules)
- The first granule to write is aligned (or aggregated files are aligned)

The released prototype will write fill granules and will take into consideration alignment (buckets).

The granules of the type `<type>` will be aggregated using the parameter `<number>`. If there is more than one version of the same granule, the granule with the latest version will be used for writing. The names of the destinations files will be generated on the fly when `<number>` of granules is identified for writing.

Two sets of aggregate files will be produced: aggregates with `<type>` granules and aggregates with the corresponding geolocation granules.

6. Other Considerations

This section describes other non-functional requirements for the prototype.

6.1. Source Code

The source code for the prototype will be in The HDF Group's publically accessible SVN repository. The repository can be accessed at <https://svn.hdfgroup.uiuc.edu/JPSSnagg>.

6.2. Build Configuration

The prototype will use GNU autotools (autoconf and automake) to build the executable.

6.3. Testing

Testing presents a challenge. The prototype will use a set of example files we have from Raytheon. The prototype will have an automated test suite to verify the

- Correctness of the group structure in the aggregated file
- Correctness of aggregated raw data for a specified product
- Correctness of the newly created attributes in the aggregated files
- Correctness of the generated file names

We will use inspections and will employ existing HDF5 command line utilities to test the prototype on a bigger collection of NPP files.

6.4. Documentation

We will create requirements and design documents and a Reference Manual entry for the prototype.

Revision History

<i>November 16, 2011:</i>	Version 1 circulated for comment within The HDF Group.
<i>November 17, 2011:</i>	Version 2 sent to Richard Ullman
<i>July 26, 2012</i>	Edited for language and usage.