GHRSST-PP

GODAE High Resolution Sea Surface Temperature Pilot Project

GHRSST-PP Accession Strategy at NODC
Draft Version 1
October 20, 2004
Kenneth S. Casey, NODC

GHRSST-PP Draft Accession Strategy

1. Each day tens to hundreds of files (file001.nc, file002.nc, …, etc.) are pulled/pushed from PO.DAAC. The file names follow the GDS convention (see attachments) and include L2P, L4, HR-DDS, MDB, and associated metadata (both Data Set Descriptions and File Records) and consist of all the files describing the calendar date 30 days previous. For example, on Sept 31, all the files containing observations from Sept 1 (or any older files that have since arrived at the PO.DAAC) would be transferred to NODC.
2. A script parses the file names and places the files from each Data Set (which is defined in the GDS by Processing Level, Sensor, and RDAC) into distinct Accessions. For example, Medspiration L2P AVHRR data would be placed in a different accession than either BlueLink L2P AVHRR or Medspiration L2P AATSR. Each accession represents one day of data from one particular type of GHRSST Data Set.
3. The script combines the Data Set Description (the quasi-static part of GCMD DIF format metadata, created by each RDAC for each Sensor and Processing Level) and File Records (variable part of GCMD DIF metadata, one for each data file) to create an FGDC metadata record for each Accession and puts it into the appropriate ABOUT/ directory.
4. The script would also move the Browse Images into the appropriate 0-DATA/ directory.
5. The script would move any other information (logs, anomaly reports, etc.) into the appropriate ABOUT/ directory.
6. The script would verify FGDC compliance and send an email alert to NODC_SOG (Casey, Philips, Barton) on failure.
7. Any file arriving that is attributed to an already existing accession would go into that accession's 02-VERSION. This situation could occur whenever an RDAC provides a new or corrected file after the 30-day window, or whenever an RDAC provides a completely reprocessed set of L2P files.
8. Since the accession structure is not very conducive to FTP/OPeNDAP, each night a script would parse the GHRSST-PP accessions and building a more suitable FTP/OPeNDAP directory tree using symbolic links to the actual data residing under the Accession structure.

NODC Canonical Accession Directory Structure

```
ACCESSIONnnnnnnn/
   01-VERSION/
         NODC_ReadMe.txt (document which describes this directory structure)
        DATA/
            0-DATA/
               file001.nc, file002.nc, …
               Browse_Images/
            1-DATA/
               (for any NODC-translated data files, but in GHRSST no translations will occur)
         ABOUT/
            Journal.txt
            FGDC_MetaDataRecord.txt
            Any other PO.DAAC provided info
   02-VERSION/
     …
```

Possible Issues:

1. A full copy of the data found in the 01-VERSION is replicated in 02-VERSION even if only a minor change is made to one file
2. Neither the Matchup Data Base (MDB) or High-Resolution Diagnostic Data Sets (HR-DDS) files are required to have associated metadata files, making formal archive at NODC difficult without some way of generating appropriate FGDC records.
3. Version control should be handled properly through the use of the 01-VERSION/, etc., directories, but the newer versions of the data must be confidently matched to previous versions.

The above plan requires or implies that...

1. The PO.DAAC need not worry about directory structure and would only need to push or provide NODC-pull capability to all the relevant files in a single directory.
2. The PO.DAAC would have to provide not only each file's individual metadata File Record (FR), but also its parent Data Set Description (DSD) each and every day. While the DSD will generally be static, it could change if an RDAC makes any changes to its processing so this approach would allow NODC to dynamically create appropriate FGDC records by combining the DSDs and FRs on a daily basis without any prior notification of a change to the DSD.
3. NODC would have to provide several key capabilities and be able to:
    a. Parse file names based on GDS conventions and move data files into appropriate new accessions on a daily basis.
    b. Combine DSDs and FRs relevant to each accession and convert from GCDM DIF to FGDC format.
    c. Verify both format (automated code with email alerts when problems arise) and content (random manual checks) of these new FGDC records.
    d. Generate proper ATDB entries dynamically with no human intervention aside from quality control (random manual checks, for example)
    e. Routinely regenerate a useful FTP/OPeNDAP directory structure composed of symbolic links based on the many daily accessions of GHRSST data.

Relevant Information from the GDS v1.5

A1.2.1 L2P filename convention (Pages 115 and 116)

The GDS filename convention used for L2P data products has been designed to provide useful information in an easily readable format. All L2P data product filenames are derived according to the following convention:

<Date Valid>-<L0 ID>-<Processing Centre Code>-L2P-<SST filename>[-<optional characteristic>]-Processing Model ID>.<base format>

which is described in Table A1.2.1

Table A1.2.1 L2P data product filename components

Name	Definition	Description
<Date Valid>	YYYYMMDD	Refers to the date for which this particular data set is valid for.
<L0 ID>	Defined in Appendix A2 Table A2.2	Data set name
<Processing Centre Code>	Defined in Appendix A2 Table A2.1	Processing centre code

<SST filename>	Native to SST filename	Filename of input SST data file as given by data provider
<optional characteristic>	string	Free field to distinguish ambiguous cases (such as ascending/descending pass when contained into a single L2 file)
<processing model ID>	v*nn* (where *nn* is the GDS version number, e.g., 01	Version number of the GDS system used to process the data file
<base format>	nc	Generic file format (nc=netCDF)

The valid date component of the filename forms the first part of the string so that data can be easily sorted by date. For example:

20030621-AVHRR16_L-AUST-L2P-LAC20030621A7SST-v01.nc

Refers to a data set that it is valid for 21st June 2003 (20030621), the source data is AVHRR NOAA 16 LAC (AVHRR16_L) that was generated at the Australian RDAC (AUST), it is a L2P data product (L2P), it is based on an input SST file called LAC20030621A7SST that was generated using the GDS version 1 (v01) and is formatted as a netCDF file (.nc).

A1.3.1 L4 product filename convention (page 131)

The GDS filename convention used for L4 data products has been designed to provide useful information in an easily readable format. All L4 data product filenames are derived according to the following convention:

> **<Date Valid>-<Processing Centre Code>-L4<Product type>-<Area>-<Processing Model ID>.<base format>**

which is defined in Table A1.3.1.

Table A1.3.1. L4 analysed data product filename components.

Name	Definition	Description
<Processing Centre Code>	Refer to Appendix A2 Table A2.1	Processing centre code
<Area>	Table A1.3.2	The area covered by the L4 product
<Date Valid>	YYYYMMDD	Refers to the date for which this particular data set

<product type>	LRfnd=low resolution,
	UHfnd=ultra-high resolution
		Resolution of analysed foundation SST (fnd) data

<processing model ID>	v*nn* (where *nn* is the GDS version number, e.g., 01	Version number of the GDS system used to process the data file
<base format>	Nc	Generic file format (nc=netCDF)

For example:

```
20040621-EUR-L4UHfnd-MED-v01.nc
```

Refers to a data set valid on 21st June 2004 (20040621) generated at the European RDAC (EUR), the data is an estimate of the foundation SST at ultra-high resolution(L4UHfnd) covering the Mediterranean area (MED), it was generated using GDS version 1 (v01) and is formatted as a GHRSST-PP netCDF file (.nc)

Table A1.3.2. L4 data product filename area code definitions. (Rev 1, 26/02/2004)

Code	Definition	Description
GLOB	90°S to 90°N and from 180°W - 180°E	Global coverer age data sets
MED	30°N to 46°N and from 6°W to 36.5°E	Mediterranean sea area
EURDAC	70°S to 90°N and from 100°W to 45°E	European RDAC area served by the ESA Medspiration project
NORSEA	48°N to 75°N and from 12°W to 70°E	Nordic Seas area
BLKSEA	40°N to 48°N and from 27°E to 42°E	Black Sea area
...		

A4.2 MDB file naming convention (page 171)

An MDB filename shall comply to the following file naming convention:

```
MDB-<Creation Date>-<Satellite file name>.<base format>
```

Table A4.2.1 Filename convention components for GHRSST-PP MDB data files.

Name	Definition	Description
<Creation Date>	yyyymmddThhmmssZ	The creation date of the MDB records contained in this file (several MDB files may be created successively for the same satellite file since some in situ observations will be delivered lately). Time specified in UTC
<satellite filename>	Related satellite filename (without the format extension)	The name of the data file from which the satellite pixels of the MDB records are extracted.
<base format>	xml	Generic file format

For example

```
MDB-20040227T123458Z-EUR-L4UHfnd-MED-v01.xml
```

Would refer to a MDB data file created on the 22nd February 2004 at 12:34:58 UTC and contains in situ data matched to EUR-SSTUHfnd-MED data produced by the GDS v1.0.

A5.2 HR-DDS data granule file format and filename convention (pages 184 and 185)

A5.2.1 HR-DDS granule filename convention

HR-DDS filenames will conform to the following format

`HRDDS_<Creation Date>_<RDAC code>_<data set>_<short name>_<long name>.<format>`

Table A5.2.1 Filename convention components for GHRSST-PP HR-DDS data files.

Name	Definition	Description
<Creation Date>	yyyymmddThhmmssZ yyyy=year mm=month (1-12) dd=day of month (1-31) T=time string identifier hh=hour of the day (0-23) mm=minute of the hour (0-59) ss=second of the minute (0-59) Z=indicates time specification is UTC.	The starting date/time of the data contained in this file. Time specified in UTC
<RDAC Code>	Table A2.1	The name of the RDAC/GDAC centre producing the HR-DDS data
<data set >	Table A2.2	GDS data set code name
<short name>	Table A5.4.1	Short name of the HR-DDS area covered by the data in this data file
<long name>	Table A5.4.1	Long name of the HR-DDS area covered by the data in this data file
<format>	Format of data contained within this data file	.nc for a netCDF data file and .png as a quicklook image file

e.g.,
`HRDDS_20040228T002133Z_EUR_ ATS_NR__2P_SST_GHR025_GHRSST-Norfolk-Island.nc`

would refer to a GDS HR-DDS data file containing data collected on the 28th February 2004 starting at 00:21:33 UTC. The HR-DDS data file was generated by the European RDAC and contains ENVISAT AATSR NR__2P 1km SST data over the HR-DDS site 'GHRSST-Norfolk-Island' and is formatted as a netCDF file.

A MMR DSD will be registered at the GHRSST-PP MMR for each HR-DDS **site** using the **HRDDS long name** that is described in Appendix A5.4 following the procedure described in Appendix A6. A MMR_FR will be generated for each HR-DDS granule and registered at the MMR following the procedure described in Appendix A6.4. Note that this arrangement is different form other data sets within the GDS where MMR_FR point to data files of the same generic type; for the HR-DDS the MMR_DSD will describe the HR-DDS site and the MMR_FR will point to many different types of GDS data products. HR-DDS filenames have been structured so that users may refer to the DSD describing the generic L2P data sets from which HR_DDS L2P data granules have been extracted and reformatted.

A6.3.1 MMR_DSD metadata record format (pages 199 and 200)

A6.3.1.1 MMR_DSD file name convention

The MMR-DSD files for L2P products shall be named as follows:

```
DSD-<L0 ID>-<Processing Centre Code>-L2P-<L2 product type>-<Processing Model
ID>.xml
```

Refer to Table A1.2.1 (L2P file naming conventions) for the meaning of each code. For example:

```
DSD-AVHRR16_L-EUR-L2P-LAC-v01.xml
```

The MMR-DSD files for L4/UHR products shall be named as follows:

```
DSD-<Processing Centre Code>-L4<Product type>-<Area>-<Processing Model ID>.xml
```

Refer to Table A1.3.1 (UHR/L4 file naming conventions) for the meaning of each code. For example:

```
DSD-EUR-L4UHfnd-MED-v01.xml
```

A6.3.2 MMR_FR metadata record format (page 212)

A6.3.2.1 MMR_FR file name convention

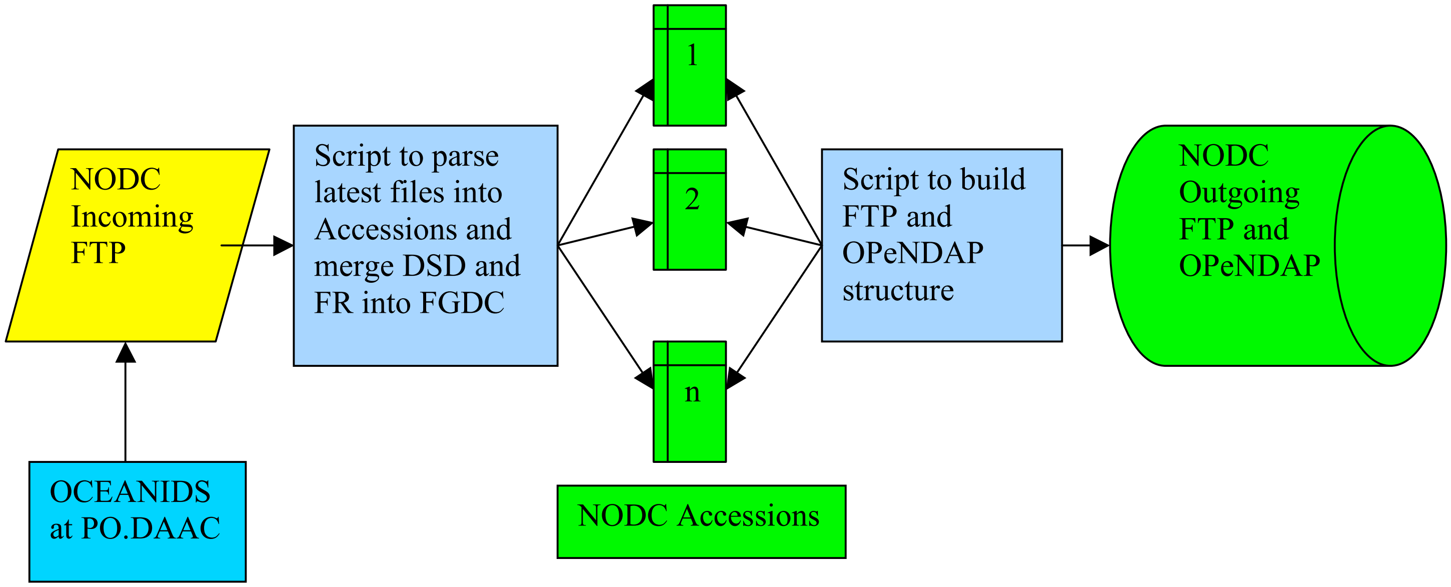
MMR-FR files for L2P and L4/UHR products shall be named as:

```
FR-<filename>.xml
```

Where `<filename>` is the name of the related L2P or UHR/L4 data file without its format extension. For example:

```
FR-20030621-EUR-L4UHfnd-MED-v01.xml
```

Conceptual Data Flow for NODC GHRSST-PP Accession Strategy


NODC Incoming FTP
Script to parse latest files into Accessions and merge DSD and FR into FGDC
1
2
n
Script to build FTP and OPeNDAP structure
NODC Outgoing FTP and OPeNDAP
OCEANIDS at PO.DAAC
NODC Accessions