Establishment of the GHRSST-PP Long Term Stewardship and Reanalysis Facility: Phase 1 – NODC Archive

Meeting Report for GHRSST-PP Seminar and Technical Meetings
Wednesday, October 20, 2004
NOAA National Oceanographic Data Center, Silver Spring, MD

Summary:
On Wednesday, October 20, 2004, numerous staff and management from the NOAA National Oceanographic Data Center (NODC) and the NASA JPL Physical Oceanography Distributed Active Archive Center (PO.DAAC) met to discuss implementation of the GODAE High-Resolution SST Pilot Project (GHRSST-PP) Long Term Stewardship and Reanalysis Facility (LTSRF) at NODC. Pat Liggett (PO.DAAC Manager), Robert Benada (Deputy Manager), Jorge Vazquez (PO.DAAC GHRSST project lead), Ed Armstrong, Andy Bingham, and Chris Finch (PO.DAAC GHRSST technical area leads) were in attendance and other PO.DAAC personnel attended via telecon as well. Many issues were discussed and an overview seminar was given as part of the NODC Science Seminar Discussion Series. Eric Lindstrom, NASA Physical Oceanography Program Manager, and Stan Wilson from NOAA attended the seminar and expressed support for the NOAA-NASA collaboration. Martha Maiden from NASA HQ attended via telecon (but had to sign off early due to a fire alarm in her building!). The working meetings following the seminar resolved several key issues including the data transmission strategy, NODC accessioning plan, and metadata conversion. Significant discussions were also held in the areas of unified user interfaces and user services collaboration. Information Management and Support Division (IMSD) Chief Parmesh Dwivedi attended the seminar and first technical session and Eric Ogata and John Relph from IMSD attended most of the day. Their input and contributions were critical and highly valued. Don Collins from the NODC Data Access Group attended most of the meetings and was extremely helpful as well. Sheri Phillips, Andrew Barton, and Ken Casey from the NODC Satellite Group were in attendance throughout the meeting. Several action items arose from the meeting, including the development of a section in the GHRSST Data Management Plan on the NODC-PO.DAAC interface and follow up on remaining technical issues regarding metadata and user services coordination.

Agenda:
The seminar and meetings were held on the 4th floor of SSMC3, in the NODC Conference Room (room 4817) and the IG Conference Room (room 4420). The agenda below was followed:

Time	Topic	Location
9:30 am – 10:00 am	Coffee and Doughnuts	NODC Conference Room
10:00 am – 11:00 am	Overview Seminar	NODC Conference Room
11:00 am – 12:00 pm	Data Transmission and Ingest	NODC Conference Room
12:00 am – 12:30 pm	Metadata Part I	NODC Conference Room
12:30 pm – 1:30 pm	Lunch	Restaurants/Cafeteria
1:30 pm – 3:00 pm	Metadata Part II	IG Conference Room
3:00 pm – 4:00 pm	User Services	NODC Conference Room

Overview Seminar:

Ken Casey from NODC and Jorge Vazquez from the PO.DAAC combined to give an overview seminar on the joint NOAA-NASA GHRSST data management activities. Their presentation is available from: http://www.nodc.noaa.gov/sog/ghrsst/GHRSSTNOAANASA041020.ppt Approximately 30 people attended the seminar including NOAA's Stan Wilson and NASA's Eric Lindstrom and Martha Maiden (via telecon, until a fire alarm cut her presence short). The seminar covered the GHRSST global/regional task sharing approach, GHRSST products and users, the science and technical challenges of GHRSST, and many details of the data management system under development and to be established cooperatively between the NOAA NODC and the NASA PO.DAAC. Both Stan Wilson and Eric Lindstrom expressed strong support for this joint NOAA/NASA collaboration.

Technical Meetings:

Following the seminar a series of technical meetings focusing on data transmission, ingest, accessioning, metadata, and user services were held. In attendance during these meetings (and the earlier seminar) were: Pat Liggett (PO.DAAC Manager), Robert Benada (Deputy Manager), Jorge Vazquez (PO.DAAC GHRSST project lead), Ed Armstrong, Andy Bingham, and Chris Finch (PO.DAAC GHRSST technical area leads). Rosanna Sumagaysay and Sue Heinz from the PO.DAAC also took part via telecon. NODC Information Management and Support Division (IMSD) Chief Parmesh Dwivedi attended the seminar and first technical session and Eric Ogata and John Relph from IMSD attended most of the day. Their input and contributions were critical and highly valued. Don Collins from the NODC Data Access Group attended most of the meetings and was extremely helpful as well. Sheri Phillips, Andrew Barton, and Ken Casey from the NODC Satellite Group were in attendance throughout the meeting. A summary of the agreements in each of the technical areas is given below:

1. Data Transmission: It was agreed that a data-pull from PO.DAAC to NODC using *ftp* or similar approach based on MD5 checksum tests and tables of contents would be the preferred method for NODC to acquire the data from the PO.DAAC, after a 30-day delay.
2. Data Ingest and Accessioning: A draft NODC GHRSST Accessioning Strategy was agreed on as a basis on which to manage the data within NODC's archive systems. This strategy is based on file name parsing using the file naming conventions for data and metadata in the GHRSST Data Specification document (GDS, Version 1, Revision 1.5). Data from a given type (L4, L2P, HR-DDS, etc) sensor (AVHRR, AATSR, etc.), RDAC, (MISST, Medspiration, etc) and day (January 1, 2005, for example) will be archived as a distinct accession. This approach seamlessly manages new versions or any changes to the incoming data sets without requiring any advance notification. On a routine, perhaps nightly, basis the GHRSST Accessions will be parsed and a more user-friendly FTP and OPeNDAP structure will be built using symbolic links and made available to the NODC/PO.DAAC GHRSST user interfaces. This accessioning plan is available at: http://www.nodc.noaa.gov/sog/ghrsst/GHRSST_AccessionStructure.pdf
3. Metadata: Since the GDS specifies that GHRSST products will be produced with GCMD DIF format metadata, separated into a static portion (Data Set Description, DSD) and a dynamic portion (File Record, FR), and since NODC is required to provide FGDC formatted metadata, several items regarding the conversion between the two were discussed. The basic idea of converting the last available DSD and any relevant FRs into a single FGDC record per accession was agreed upon in principle. It was also felt that the tools to handle this are also available, so NODC personnel will be identifying these tools

and insuring the there will be sufficient information in a GCMD DIF to create a minimally acceptable FGDC record. Some issues were left to action items including understanding the metadata requirements for the Matchup Database (MDB) records and whether they would be sufficient for this proposed system.

4. Users Services: Sue Heinz presented an overview of the proposed User Services approach for management GHRSST user interactions between NODC and the PO.DAAC. Issues discussed included a unified branding approach and the seamless presentation of data, whether it resides physically at the PO.DAAC (less then 30 days old) or at NODC (30 days and older).

Action Items:

1. Pat Liggett: Investigate whether any formal MOU/MOA's are required between NOAA and NASA for this collaborative activity.
2. Andy Bingham: Develop section of GHRSST Data Management plan documenting the details of the NODC-PO.DAAC interface and distribute for comment.
3. Chris Finch and John Relph: Verify network is capable of the required data transfer rates.
4. Andy Bingham and John Relph: Finalize details of data transmission (hostnames, scheduling, etc.)
5. Ed Armstrong: Establish file name convention and appropriate frequency for MDB file dumps to be saved as NODC accessions.
6. Ed Armstrong and Andy Bingham: Understand Master Metadata Repository (MMR) transfers and what to include in a given accession.
7. Ed Armstrong: Follow up with Craig Donlon (Chair of GHRSST-PP) and Dave Poulter on latest developments in Europe regarding High Resolution Diagnostic Data sets (HR-DDS) and Matchup Database (MDB):w.
8. Andy Bingham: Identify which quality assurance and metadata checks are being done in PO.DAAC OCEANIDS systems and coordinate with Ken Casey to define a set of necessary checks to insure data and metadata quality.
9. Ken Casey: Establish draft of FTP/OPeNDAP structure to be shared jointly by NODC and PO.DAAC and distribute for comment and approval.
10. Sheri Phillips: Complete investigation of GCMD DIF to FGDC conversion issues, insuring minimally compliant FGDC records can be generated.
11. Ken Casey: Write meeting report, circulate for comments, and provide to Craig Donlon for posting to GHRSST web site and sharing with GHRSST Science Team.
12. Ken Casey: Coordinate follow-up meetings via telecon and perhaps in person at upcoming Fall AGU meeting in December.

Establishment of the GHRSST-PP Long Term Stewardship and Reanalysis Facility
Phase 1 – NODC Archive
Joint NOAA NODC and NASA JPL PO.DAAC Meeting Report
October 20, 2004
Silver Spring MD 20910 USA

Drafted by Kenneth S. Casey, NOAA NODC
Kenneth.Casey@noaa.gov
301-713-3272 x133