Implementation Plan for the GODAE High Resolution Sea Surface Temperature Pilot Project Reanalysis System

GHRSST-PP Document Reference xx/Version 1.0
07 September 2005
Prepared by Kenneth S. Casey for the RAN-TAG

Executive Summary
The GODAE High Resolution Sea Surface Temperature Pilot Project (GHRSST-PP) has been established to provide international focus and coordination for the development of a new generation of global, multi-sensor, high-resolution (better than 10 km), sea surface temperature (SST) products provided in near real time (6 hourly). Its primary aim is to oversee the development, timely delivery, assembly and processing of high-quality, global scale, SST products at a fine spatial and temporal resolution, for the diverse needs of GODAE and the wider scientific community.

The GHRSST-PP Reanalysis Technical Advisory Group (RAN-TAG, Annex 1) is the formal GHRSST-PP body that is responsible for the scientific and operational methods and algorithms used to generate delayed-mode GHRSST-PP data products suitable for use as climate data records (CDRs). The terms of reference, goals, and product target characteristics were formalized by the GHRSST-PP Science Team in the RAN-TAG's First Report (GHRSST/19).

This document contains the RAN Implementation Plan, which describes the system and components needed to deliver RAN products which meet the goals stated in the First Report. The RAN system will centered at the Long Term Stewardship and Reanalysis Facility located the NOAA National Oceanographic Data Center (NODC), but will also consist of components distributed among the GHRSST-PP Global Data Analysis Centre (GDAC) and Regional Data Assembly Centre (RDAC) facilities that exist within the global/regional task sharing model that forms the basis of the GHRSST-PP Data Processing Specification (GDS, GHRSST/17). Further information, including all reference material, can be found at the
GHRSST-PP international Project Office web site: http://www.ghrsst-pp.org.

Document Change Record

Author	Modification	Issue	Revision	Date
K.S. Casey	Original	DRAFT	0.1	20 July 2004
K.S. Casey	Updated with recommendations from GHRSST-5 meeting in Townsville, results of the PO.DAAC-NODC meeting held in October 2004 at NODC, and developments made during establishment of the LTSRF during early 2005.	For ST Review	0.9	11 May 2005
K.S. Casey	Updated with new RAN-TAG membership and comments from GHRSST-6 meeting in Exeter, UK.		1.0	7 September 2005

Table of Contents

Introduction

Experience with the generation and application of real-time satellite sea surface temperature (SST) data sets indicates a clear need for sustained, long-term reprocessing efforts to optimize product accuracy and temporal stability. Based on this experience, the GHRSST-PP Science Team initiated a Reanalysis (RAN) program whose goals are to produce delayed-mode products of higher accuracy and consistency than the real-time SSTs. By taking advantage of additional delayed mode data streams that cannot be used by the operational real time system, scientific developments over time, and by conducting routine reprocessing of the entire time series using a consistent set of techniques, the RAN system will enable a wide range of additional applications of GHRSST-PP products. For example, the reprocessed RAN products will be suitable for use in climate studies as "climate data records" (CDRs), an important concept in environmental data stewardship which dictates long-term accuracy and consistency (*e.g.* NRC, 2000).

The GHRSST-PP Reanalysis Technical Advisory Group (RAN-TAG, Annex 1) is the formal GHRSST-PP body that is responsible for the scientific and operational methods and algorithms used to generate RAN products. This document contains the RAN Implementation Plan, which describes the system and components needed to deliver RAN products which meet the goals stated in the RAN-TAG First Report (GHRSST/19). The RAN system will centered at the Long Term Stewardship and Reanalysis Facility located the NOAA National Oceanographic Data Center (NODC), but will also consist of components distributed among the GHRSST-PP Global Data Analysis Centre (GDAC) and Regional Data Assembly Centre (RDAC) facilities that exist within the global/regional task sharing model that forms the basis of the GHRSST-PP Data Processing Specification (GDS, GHRSST/17). After this introduction, a description of the overall RAN system will be presented followed by descriptions of each required component.

Overall System Architecture

The overall system is composed of distributed components, following the overall GHRSST-PP architecture as defined in the GDS. In this distributed system, careful attention must be paid to each of the component interfaces. The RAN system will conform to the GDS and other relevant GHRSST-PP interface documents whenever possible.

The components of the RAN system are illustrated in **Figure 1** and include:

1. Real-time input data streams
2. Delayed-mode input data streams
3. Reanalysis Community
4. Long Term Stewardship and Reanalysis Facility (LTSRF)
    a. Ingest
    b. Archive
    c. Access
    d. Application
    e. User Services
    f. Reanalysis Functions

Of these components, it is expected that the Long Term Stewardship and Reanalysis Facility

(LTSRF) will play a crucial role since it interfaces with the diverse input data streams through the GDAC at the NASA JPL Physical Oceanography Distributed Active Archive Center (PO.DAAC) to achieve end-to-end functionality for the system, including the required computational capability for reprocessing and reanalysis. The NOAA National Oceanographic Data Center (NODC) is developing the capacity to serve as the LTSRF. More details on these components and their interfaces are given below.

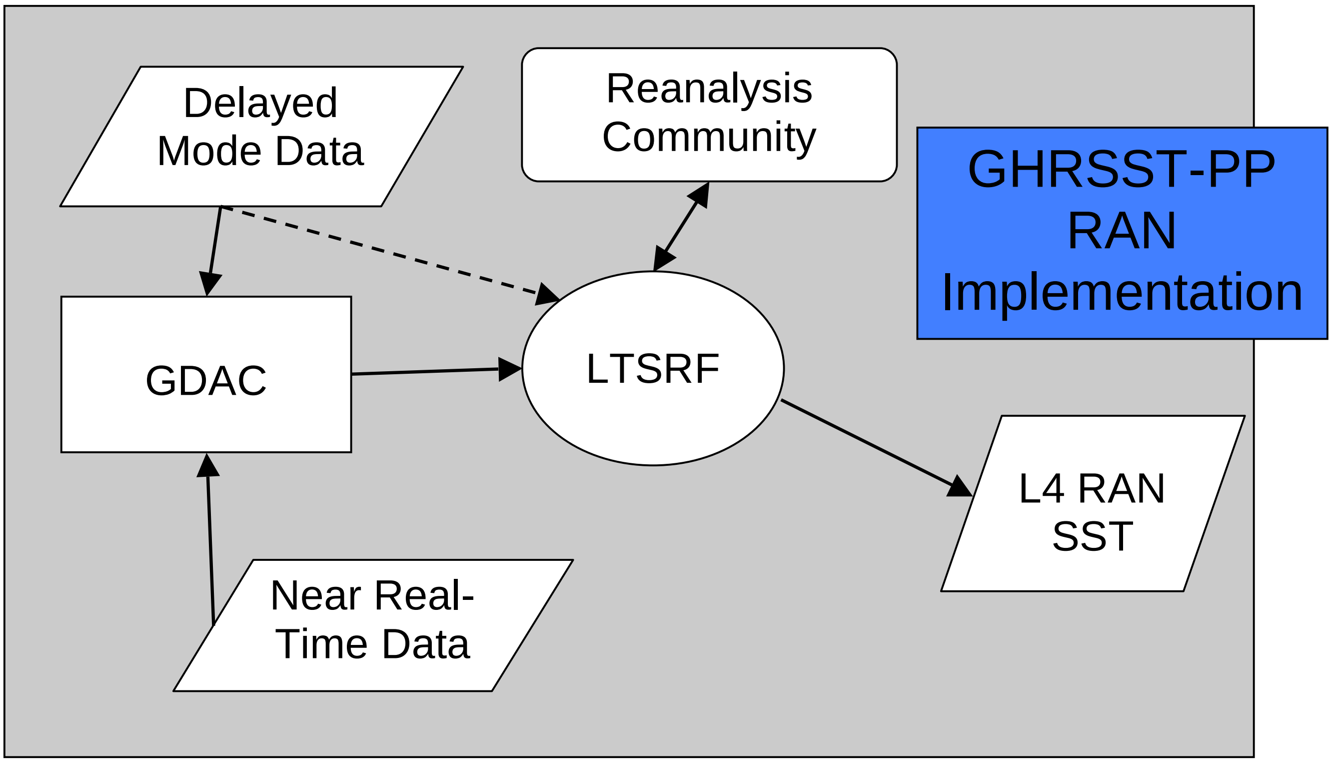

Figure 1: Schematic diagram of GHRSST-PP RAN implementation, illustrating the core components of the system.

Real-time Input Data Streams

Description

One of the fundamental inputs to the GHRSST RAN are the real time, Level 2 Preprocessed (L2P) SST data sets. In association with these geophysical records are the Matchup Data Base (MDB), the High Resolution Diagnostic Data Sets (HR-DDS, GHRSST-PP/14), and Master Metadata Repository (MMR). Each of these data sets arrives in near real time to the PO.DAAC, which serves as the GDAC for the real time GHRSST products. It is at the GDAC that the various inputs from all the RDACs are brought together in a unified framework. This framework then enables the real time production of Level 4 (L4) analyzed output fields. The GHRSST-PP Data Management Plan (GHRSST/RD-6) defines the interfaces between the GDAC and the RDACs which provide the L2P and associated data.

<u>Interfaces</u>

After a period of 30 days from observation, the real time data streams will be sent to the LTSRF at NODC for stewardship in perpetuity. In the context of the GHRSST RAN system, these real time data streams are defined to include any late-arriving or corrected L2P and associated data that is ingested at the GDAC within the 30 day window. The purpose of the delayed transfer from GDAC to LTSRF is to provide the opportunity for the GDAC to ingest these late arriving or corrected data streams before transferring to the LTSRF. Experience with other operational systems indicates that this delay will substantially simplify the GDAC-LTSRF interface, since corrections and late-arriving data are commonplace. The data transfer from the PO.DAAC GDAC to the NODC LTSRF is through an automated FTP-pull initiated on a routine basis by NODC. Data not arriving at the GDAC within 30 days are handled in the GHRSST RAN system as delayed mode data streams, as described in the following section.

Delayed Mode Input Data Streams

<u>Description</u>

In the context of GHRSST RAN, delayed mode data streams are those which were not ingested by the GDAC within the allotted 30 day window. These delayed mode data may include very late arriving L2P files, substantially delayed in situ matchup data in the MDB, updated metadata in the MMR, improved HR-DDS data, and even entirely reprocessed or corrected L2P products streams produced by an RDAC. These latter groups of reprocessed L2P are expected to form the baseline inputs for the GHRSST RAN and include such datasets as AVHRR Pathfinder and new developments in Europe regarding (A)ATSR reprocessings. Delayed mode data are particularly challenging to manage properly in the GHRSST RAN system since they can become available to the LTSRF from the GDAC or directly to the LTSRF from an outside source. To minimize the challenge, it is expected that all delayed mode data streams originating from RDACs will also be ingested first through the GDAC. Situations where this approach is not possible or feasible will be handled on a case-by-case basis. Careful management of version levels within the LTSRF is critical.

<u>Interfaces</u>

A careful definition of the interfaces between the possible sources of delayed mode data and the LTSRF is critical to enable efficient use of the data and accurate inclusion with the rest of the GHRSST-PP database. The first interface connects the LTSRF and GDAC. This interface is already well defined since the GDAC provides all the real time data streams to the LTSRF as described in the previous section. The second interface connects the LTSRF with other delayed mode sources of data such as related science efforts, buoy programs, or other groups originating data useful to the RAN.

Reanalysis Community

<u>Description</u>

The most fundamental requirement in the production of climate data records is acceptance by the community which it serves (e.g., NRC 2004). Developing this community and its acceptance is an iterative process, as exemplified by other successful satellite reprocessings like the AVHRR Pathfinder SST program and the SeaWiFS Ocean Color system. Both of these successful reprocessing schemes developed community consensus through numerous reprocessing–user feedback–reprocessing cycles.

Interfaces

To develop a reanalysis community and its acceptance of GHRSST-PP RAN products as CDRs, clear interfaces with the user community must be maintained. Traditionally, these interfaces have taken place at the point of development and distribution of the satellite-based CDR. In the GHRSST-PP RAN context, the community interface must work well with the LTSRF. Users must be given the opportunity to provide input on algorithm development, on the quality of the data, and must be able to see their comments addressed seriously. For example, in the recent Pathfinder Version 5.0 SST reprocessing at NODC and UMiami, users noticed problems with the land mask off South America and a subtle problem occurring every 18 lines when examining time series of gradient data derived from the SST fields. These problems are now being fixed as part of the next reprocessing (Version 6) because of the direct and close connection between users and the data providers and developers.

Long Term Stewardship And Reanalysis Facility

Description

The LTSRF will form the central hub of the distributed GHRSST-PP RAN system. NODC is currently developing the LTSRF capabilities. As the name LTSRF implies, complete stewardship, not simple archive in the traditional sense, is required to enable a successful RAN system. Six broad activities define the overall stewardship responsibilities of the LTSRF:

1. **Ingest**: Ingest involves the receipt, verification, and proper cataloging, via appropriate file-level and collection-level metadata, of the input data streams, defined above as including both real-time and delayed-mode data.
2. **Archive**: Archive activities include offsite backup, media migration, and validation of stored data.
3. **Access**: These activities include the critical role of providing data and metadata to a both a diverse user community and to the LTSRF itself, following relevant standards. It is expected that all data will be provided in a free and open manner via the internet. Fees for media distribution (CD, DVD, etc.) would be limited to cost of production only.
4. **Application**: To provide proper long term stewardship of any data, it is important for data providers/developers to use and test that data in a variety of applications with feedback on the data quality. In doing so, the data providers/developers become much more knowledge about the data, identify weakness sooner, and more quickly validate the use of the data in new application areas.
5. **User Services**: The User Services component of the LTSRF not only provides the standard user assistance with questions about the data, but also serves as the vital feedback loop to receive user input on problems, new applications, and directions for future improvements. This component will be primarily achieved through user services

support of the GDAC at the PO.DAAC.

6. **Reanalysis Functions**: The LTSRF must be capable of reprocessing the entire GHRSST-PP collection at a sufficiently rapid pace. The required rate is dependent on the length of the record to be reprocessed and the desired overall reprocessing time. The length of record to be reprocessed will grow gradually as time passes and the primary input L2P records grow in length, but may also grow rapidly at discrete intervals if the capability to extend the RAN period backward in time to pre-GHRSST-PP years is developed. The desired overall reprocessing time will remain fairly fixed and will likely be defined as around three months. By conducting the entire reprocessing within this window of time, sufficient time will exist for focusing on the important tasks of validating the reprocessed L4 data, interfacing with the community, and making algorithm improvements for the next round of reprocessing. As an example, to reprocess one year of data within three months would require a rate of only 4 days of data processed per day of real time. As the number of years to reprocess grows, this rate may substantially increase. The AVHRR Pathfinder reprocessing effort, for example, reprocesses 20 years of data and would have to be capable of reprocessing about 60 days of data per day of real time to complete the entire record within 3 months (it presently can not quite achieve that rate).

<u>Interfaces</u>

The LTSRF must be fully interfaced with all parts of the RAN system to insure a successful GHRSST-PP RAN capability. The first of these interfaces is between the LTSRF and the PO.DAAC, which serves as the GDAC where the real-time data streams are integrated. This interface has already been discussed in the sections on Real Time Data Streams and Delayed Mode Data Streams. Another critical interface is between the LTSRF and the community of users and scientists interested in using and developing the RAN data. This interface would be maintained through the User Services component, actively engaging the community with a better, more useful data set resulting. The LTSRF will also provide interfaces to external reprocessing efforts, supporting the idea that multiple, independent reprocessing efforts can provide important insights on climate data record generation (NRC, 2004).

References

GHRSST-PP/14. Draft GHRSST-PP High Resolution Diagnostic Data Set Implementation Plan, available at http://www.ghrsst-pp.org, 2004.

GHRSST-PP/17. The GHRSST-PP Data Processing Specification v1.0, revision 1.5, available at http://www.ghrsst-pp.org, 2004.

GHRSST-PP/19. First Report of the GHRSST-PP Reanalysis Technical Advisory Group, available at http://www.ghrsst-pp.org, 2004.

GHRSST-PP/RD-6. First Report of the GHRSST-PP Data Management Technical Advisory Group, available at http://www.ghrsst-pp.org, 2004.

NRC (2000). Ensuring the climate record from the NPP and NPOESS meteorological satellites, NAP, 71 pp.

NRC (2004). Climate Data Records from Environmental Satellites, Interim Report. NAP, 105 pp.

GHRSST-PP Contacts

Donlon, Craig J.
Director, GHRSST-PP Project Office
Met Office
Fitzroy Road
Exeter, EX3 1PB
United Kingdom
Tel: +44 (0)1392 886622
Fax: +44 (0)1393 885681
E-mail: Craig.Donlon@metoffice.com

Smith, Neville
Chairman of GODAE
BMRC
Box 1289K
Melbourne
Vic. 3001
AUSTRALIA
Tel: +61 3 9669 4434
Fax: +61 3 9669 4660
E-mail: N.Smith@bom.gov.au

Annex 1: RAN-TAG Membership

Casey, Kenneth S. (Chairperson) NOAA/ National Oceanographic Data Center 1315 East-West Highway Silver Spring, MD 20910, USA Phone: +01 301-713-3272 x133 Fax: +01 301-713-3300 E-mail: Kenneth.Casey@noaa.gov	Emery, William University of Colorado CB 431 Boulder, CO 80309-0431, USA Phone: +01 303-492-8591 Fax: +01 303-492-2825 E-mail: William.Emery@colorado.edu
Armstrong, Edward Jet Propulsion Laboratory/California Institute of Technology 300/323 4800 Oak Grove Drive Pasadena, CA 91109, USA Phone: +01 818-354-6980 Fax: +01 818-393-6720 E-mail: ed@seastar.jpl.nasa.gov	Reynolds, Richard W. NOAA/NESDIS National Climatic Data Center 151 Patton Avenue Asheville, NC 20881, USA Phone: +01: (828) 271-4302 Fax: +01 (828) 271-4328 E-mail: Richard.W.Reynolds@noaa.gov
Vazquez, Jorge Jet Propulsion Laboratory/California Institute of Technology 300/323 4800 Oak Grove Drive Pasadena, CA 91109, USA Phone: +01 818-354-6980 Fax: +01 818-393-6720 E-mail: jv@pacific.jpl.nasa.gov	Wick, Gary A. NOAA Environmental Technology Laboratory R/ET6 325 Broadway Boulder, CO 80305, USA Phone: +01 303-497-6322 Fax: +01 303-497-6181 E-mail: Gary.A.Wick@noaa.gov
Rayner, Nick UK Met Office Hadley Centre for Climate Prediction and Research London Road, Bracknell, RG12 2SY, UK Phone: +44 1344 854063 Fax: +44 1344 854898 E-mail: mailto:nick.rayner@metoffice.gov.uk	Donlon, Craig J. Director, GHRSST-PP Project Office, Met Office Fitzroy Road Exeter, EX3 1PB United Kingdom Tel: +44 (0)1392 886622 Fax: +44 (0)1393 885681 E-mail: mailto:Craig.Donlon@metoffice.gov.uk
Kawamura, Hiroshi Center for Atmospheric and Oceanic Studies Graduate School of Science Tohoku University Sendai 980-8578 Japan Phone:+81 22-217-6745 Fax:+81 22-217-6748 Email: kamu@ocean.caos.tohoku.ac.jp	Murray, Jo Space Science and Technology Department CCLRC Rutherford Appleton Laboratory Chilton, Didcot Oxfordshire, OX11 0QX UK Tel: +44 1235 446507 Fax: +44 1235 445848 E-mail: Jo.Murray@rl.ac.uk
Larnicol, Gilles CLS Space Oceanography Division 8-10 rue Hermes Parc Technologique du Canal 31526 Ramonville Saint-Agne, France Tel:+33 5 6139 3737 Fax: +33 5 617 51014 Email: glarnicol@cls.fr	