Climate Data Record (CDR) Program

Transitioning CDRs from Research to Operations

CDR Program Document Number: CDRP-PLAN-0017
Originator Document Number: N/A
Version 1.0 / July 26, 2011

TABLE of CONTENTS

1. CDR RESEARCH-TO-OPERATIONS OVERVIEW ... 4
1.1 Introduction ... 4
1.2 CDR R2O Process ... 5
1.3 Definition of IOC ... 6

2. GENERAL ROLES AND RESPONSIBILITIES ... 6
2.1 CDR Program Office ... 6
2.2 Principal Investigator ... 6
2.3 Products Branch ... 6
2.4 Archive Branch ... 7
2.5 Operations Branch ... 7
2.6 Information Technology Branch ... 7
2.7 Access Branch ... 7

3. CDR TRANSITION PROCESS ... 8
3.1 Assessment Phase ... 8
3.1.1 Initial PI Assessment Package ... 8
3.1.2 CDR Assessment ... 8
3.2 Preparation Phase ... 8
3.2.1 Source Code ... 8
3.2.2 Documents ... 9
3.2.3 Data ... 9
3.2.4 Submission Agreement ... 9
3.3 Transfer Phase ... 9
3.3.1 Send Samples ... 9
3.3.2 Readiness Review ... 10
3.3.3 Main Data Transfer ... 10
3.3.4 Version Controlling Source Code ... 10
3.3.5 Security Review ... 10
3.4 Verification Phase ... 10
3.5 Archival Phase ... 10
3.6 Access Phase ... 11
3.7 IOC Maintenance ... 11
3.8 IOC to FOC ... 11

APPENDICES ... 11
Appendix A – CDR Assessment Example ... 12
Appendix B – Submission Agreement Example ... 16
Appendix C – Code Header Document Example ... 22
Appendix D – Data Flow Diagram Example ... 24
Appendix E – CDR Maturity Matrix ... 25
Appendix F – IOC Checklist ... 26
Appendix G – GCOS 12 Requirements for a CDR ... 27
Appendix H – Frequently Asked Questions ... 28

LIST of FIGURES

Figure 1: CDR End-to-End Concept. 4

Figure 2: CDR R2O Process Diagram...5

LIST of TABLES

Table 1. GCOS 12 Requirements for CDRs 27

ACRONYMS and ABBREVIATIONS

Acronym or Abbreviation	Meaning
C-ATBD	Climate Algorithm Theoretical Basis Document
CDR	Climate Data Record
CDRP	Climate Data Record Program
CIR	Climate Information Record
FOC	Full Operating Capability
GCOS	Global Climate Observing System
IOC	Initial Operating Capability
IPT	Integrated Process Team
IT	Information Technology
KDP	Key Decision Point
MM	Maturity Matrix
NCDC	National Climatic Data Center
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center
PI	Principal Investigator
POC	Point of Contact
POR	Period of Record
R2O	Research-to-Operations
RSAD	Remote Sensing and Applications Division
SA	Submission Agreement
SME	Subject Matter Expert
THREDDS	Thematic Realtime Environmental Distributed Data Services

1. CDR Research-to-Operations Overview

1.1 Introduction

These guidelines define the process used by the Climate Data Record Program (CDRP) to transition a Climate Data Record (CDR) from research to Initial Operating Capability (IOC) at the National Climatic Data Center (NCDC). This process is commonly called the CDR Research-to-Operations (R2O) Transition Process. A subsequent process will further transition these products from IOC to Full Operating Capability (FOC) but is outside the scope of this document. Figure 1 is a diagram illustrating the CDR end-to-end concept. The guidelines defined in this document only pertain to the R2O subsection of Figure 1.

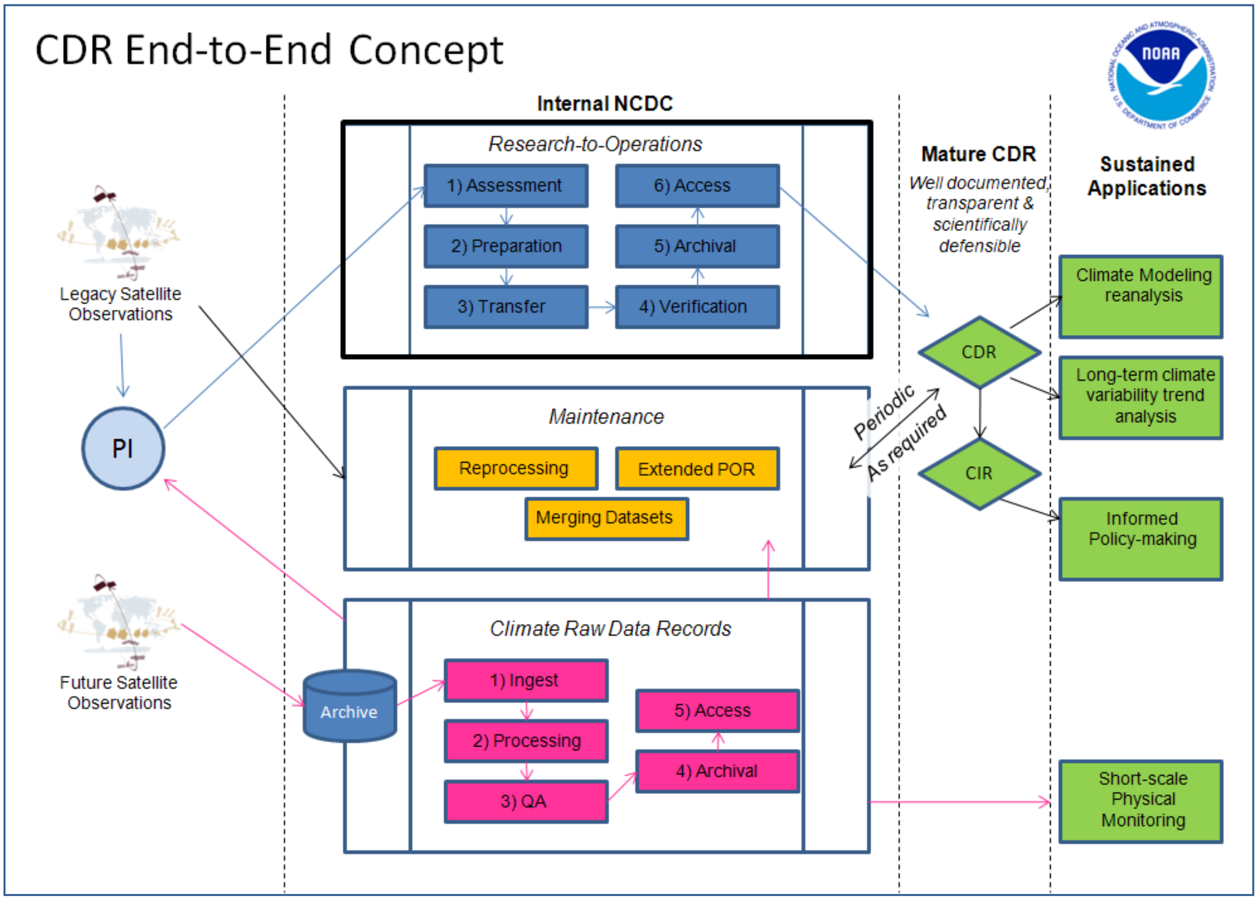


Figure 1: CDR End-to-End Concept.

For the purpose of this document, 'external' is defined as CDR algorithm software, data and supporting documentation originating from a Principal Investigator (PI) outside of NCDC as opposed to 'internal' which originates from within NCDC. Regardless of origin, these guidelines ensure all CDRs meet the Global Climate Observing System

(GCOS) standard of quality outlined in Appendix G. The maturity of an operational CDR can be described as either Initial Operating Capability (IOC) or Full Operating Capability (FOC). IOC processes describe the preliminary R2O requirements that need to be met before making a CDR accessible to the public. Acknowledging there is no 'end state' to a climate record, FOC processes describe the stewardship that maintains continuous production and cyclical improvements to a CDR. The guidelines defined in this document specifically deal only with IOC processes.

1.2 CDR R2O Process

Figure 2 illustrates the CDR R2O process at NCDC required to transition from CDR Assessment to IOC. There are six phases toward preparing a CDR for IOC: Assessment, Preparation, Transfer, Verification, Archival, and Access. The entire process is a collaborative effort touching five separate branches spanning three divisions within NCDC, as well as the PI. Successful transition of CDRs requires excellent coordination and communication among all parties. Each party needs to understand and execute their roles and responsibilities for the process to be successful.

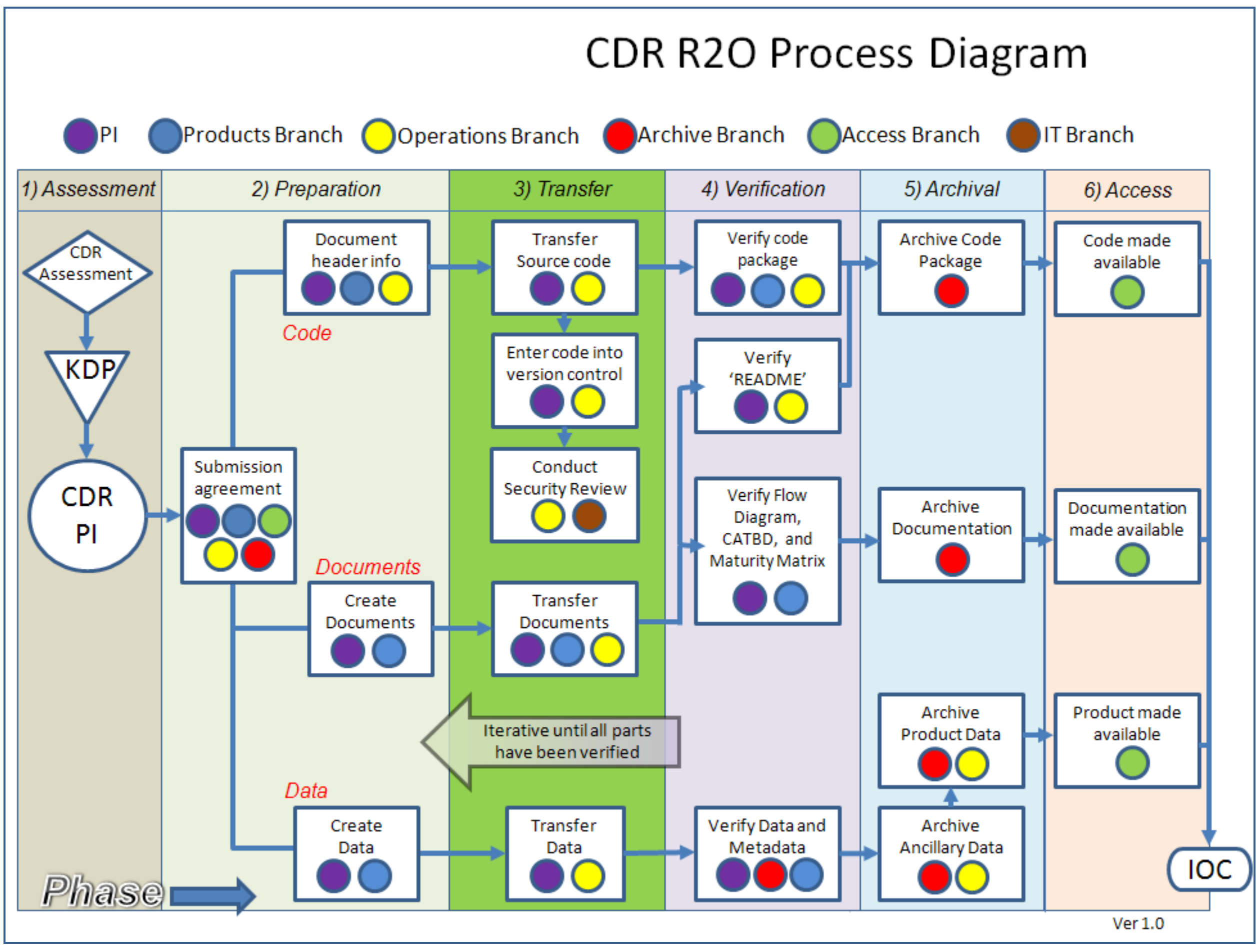


Figure 2: CDR R2O Process Diagram

1.3 Definition of IOC

A CDR is considered IOC once it has met all the source code, documentation, and data requirements; is preserved in the NCDC archive; and is made publicly accessible. The CDR is at a level of maturity such that all criteria on the Maturity Matrix in Appendix E are at level 3 or greater. Meeting these requirements ensures that the final CDR is transparent and scientifically defensible. IOC is considered the first iteration of the publicly released CDR.

2. General Roles and Responsibilities

2.1 CDR Program Office

The CDR Program Office serves a CDR planning and management function for the Remote Sensing and Applications Division (RSAD) and NCDC. The Program Office is responsible for the business affairs such as performance metrics, budgets, grants and contracts management. To maintain oversight on the multiple CDRs in transition the CDR Program Manager will appoint a R2O Project Manager to monitor the CDRs throughout the R2O transition process. The R2O Project Manager will provide weekly status reports to the CDR Program Manager and RSAD Branch Chiefs. The R2O Project Manager will also provide all parties the requirements, guidelines, and templates necessary to transition a CDR from research to IOC.

2.2 Principal Investigator

The PI is responsible for providing the source code, documentation, and data necessary to establish a CDR that is transparent and scientifically defensible. The process to achieve this is further detailed in Section 3 and the Appendices.

2.3 Products Branch

The Products Branch Chief will designate a lead scientific Point of Contact (POC) for each CDR research to operations initiative. This POC is responsible for ‘cradle to grave’ nurturing of the CDR to include initial assessment, verification of PI submissions and coordinating the intra-NCDC collaborative effort to IOC. This may require the formation of an Integrated Project Team (IPT) consisting of subject matter experts (SME) that help facilitate the transition process. The POC will provide weekly updates to the R2O Project Manager that details the transition status, future milestones, and any potential risks to completing the transition on schedule.

2.4 Archive Branch

The Archive Branch Chief will appoint a team member to facilitate the archival process for long term stewardship of the CDR source code, documentation, and data. The details of the stewardship process including metadata content and format, file naming conventions and submission packaging are captured by the POC in the submission agreement, which requires significant coordination between the PI, POC, Operations, Access and Archive Branches. The Archive Branch will participate in management and technical meetings as required to facilitate communication and successful completion of IOC.

2.5 Operations Branch

The Operations Branch Chief will appoint a team member to facilitate the transfer of data to NCDC for the archival process. The details of the transfer are identified in the submission agreement. This team member will also verify source code and Readme packages for compliance with standards, enter the source code into the NCDC version control system and request an initial security review. The Operations Branch will participate in management and technical meetings as required to facilitate communication and successful completion of IOC.

2.6 Information Technology Branch

The Information Technology (IT) Branch will conduct a security review of the source code when requested by the Operations team member. They are also responsible for IT support of the CDR systems. The IT Branch will participate in management and technical meetings as required to facilitate communication and successful completion of IOC.

2.7 Access Branch

The Access Branch will ensure that the CDR source code, documentation and data components are all made available to the public. The Access Branch will participate in management and technical meetings as required to facilitate communication and successful completion of IOC.

3. CDR Transition Process

3.1 Assessment Phase

3.1.1 Initial PI Assessment Package

In order to conduct an assessment of a potential CDR, the CDR Program Office will introduce the TM and ask the PI for an initial product package. The contents of this package should include a significant portion of the source code, draft Climate Algorithm Theoretical Basis Document (C-ATBD) and data flow diagrams along with a sample of product data. The TM will review the package as part of the CDR assessment.

3.1.2 CDR Assessment

Prior to committing significant NCDC resources to a specific CDR, a comprehensive assessment will be conducted by the POC to determine whether the CDR is mature enough for transition. This assessment will help determine the level of effort needed to meet all the IOC requirements listed in Appendix F. The POC will draft the formal recommendation to the CDR Program Manager whether the CDR is ready for IOC transition. A sample CDR assessment is shown in Appendix A. The CDR Program Manager reviews the assessment and makes the final determination to proceed with the transition to IOC or not. This is the Key Decision Point (KDP) shown in Figure 2. If the CDR is chosen for transition an estimate of the data volume and data types will be provided to the Archive Branch for planning purposes. A delivery schedule is then coordinated between the PI, POC, and R2O project manager to ensure that the multiple CDRs being transitioned are staggered. This is to prevent any potential resource conflicts that could arise from too many CDRs being in the end phases at the same time.

3.2 Preparation Phase

The preparation phase is typically the longest and most labor intensive. After the CDR is selected for transition to IOC, the CDR Program Office will inform the PI to start work on the CDR package. The POC will start drafting the submission agreement to ensure that all parts of the CDR package will be properly archived. The CDR package consists of well documented source code, supporting documentation, and the data.

3.2.1 Source Code

The PI will ensure the source code meets the minimum standards listed in the *CDR Program General Software Coding Standards* (http://www.ncdc.noaa.gov/cdr/guidelines.html). The source code will also be accompanied by a README file which provides step by step instructions on how to setup and run the source code.

3.2.2 Documents

The PI will create the following documents:

- Climate Algorithm Theoretical Basis Document (C-ATBD) using the template provided by the CDRP office (http://www.ncdc.noaa.gov/cdr/guidelines.html).
- Data Flow Diagram outlining the CDR production process (e.g. Appendix D)
- Completed Maturity Matrix (http://www.ncdc.noaa.gov/cdr/guidelines.html). In order to reach IOC all columns must be rated at level 3 or above (e.g. Appendix E)
- Code Header Document describing each unit of code by file name, purpose, author, creation date, copyright and other pertinent information required by the *CDR Program General Software Coding Standards*. The example in Appendix C was generated using robodoc.

3.2.3 Data

The PI will ensure the CDR data set is in NetCDF format, contains the required metadata described in *NetCDF Metadata Guidelines for IOC Climate Data Records* (http://www.ncdc.noaa.gov/cdr/guidelines.html) and adheres to the file naming conventions documented in the Submission Agreement (SA). The PI will also identify any unique input or ancillary data sets that are required to create the CDR so the POC can add them to the SA if needed.

3.2.4 Submission Agreement

The POC drafts the SA which defines in detail the CDR submission package for ingest and archive, data transfer mechanism and specific data access needs as well as identifying specific roles and responsibilities. Drafting the SA involves coordinating with several NCDC branches and the PI and may require the formation of an IPT to facilitate communication. The SA can be written using the SA template or the online ATRAC tool (https://www.ncdc.noaa.gov/atrac/index.html). The Archive Branch team member will review the SA for accuracy and completeness. This agreement is a living document as long as the data reside within the NCDC archive. The POC, as well as the Operations and Archive team members, will stay actively engaged with the PI during the entire submission process. An example of a CDR submission agreement is in Appendix B.

3.3 Transfer Phase

3.3.1 Send Samples

As progress is made by the PI on the source code, documents, and data, samples of each should be sent to the POC to ensure the guidelines and standards are being met. This helps the PI by ensuring that their efforts are on track or can be refocused if needed. Subsequently the Preparation, Transfer, and Verification Phases are closely related and will be iterative until all the parts have been fully verified.

3.3.2 Readiness Review

Before the PI begins the main data transfer the Operations Branch team member will ensure that the transfer protocols described in the SA are operational and that the NCDC ingest location has sufficient space to receive the CDR.

3.3.3 Main Data Transfer

The PI and Operations Branch team member will monitor the transfer of data to NCDC, and verify that it transferred correctly as described in the SA.

3.3.4 Version Controlling Source Code

With the assistance of the PI and POC, the Operations Branch team member will enter the CDR code into the NCDC version control system. Version control allows easy management of changes to documents, code and other information stored as computer files. This is also the staging area for source code requiring a security review.

3.3.5 Security Review

After the source code has been entered into the NCDC version control system, the Operations Branch team member will request an initial security review be conducted by the IT Branch. Although passing the security review is not a prerequisite for IOC it does have a twofold purpose. First, it is a way to provide feedback to the PI on any security vulnerabilities found during the initial review of their code. Second, the security review allows for future personnel resource planning when transitioning from IOC to FOC. When a CDR is selected for FOC transition, any vulnerabilities that were identified will need to be fixed before the code can be deployed in the production environment.

3.4 Verification Phase

CDR source code, documentation and data need to be verified by the POC, Operations and Archive Branch team members, to ensure the submission meets the required guidelines and standards. It is optimal for NCDC to receive test files of source code, documents or data before transfer of the final submission to identify any potential problems early in the process. Any items that fail verification should be returned to the PI for corrections.

3.5 Archival Phase

After all parts of the CDR have been verified, the Archive Branch stores the code package (source code, README, and results of the security review), documentation package (C-ATBD, Flow Diagram, Maturity Matrix, and Code Header Document), and the data (CDR product and any required input/ancillary data) in the archive as outlined in the SA. Ensuring that the source code, documentation and data are all archived with a unique version identifier satisfies Global Climate Observing System (GCOS) recommendation 7, *‘Version management of FCDRs and products, particularly in connection with improved algorithms and reprocessing’* (Ref. Appendix G).

3.6 Access Phase

The role of the Access Branch is to provide public access to the CDR source code, all documentation and data components. This is typically accomplished via the CDR web site (http://www.ncdc.noaa.gov/cdr/operationalcdrs.html). This final phase satisfies Global Climate Observing System (GCOS) recommendation 8, *‘Arrangements for access to FCDR, products, and all documentation’* (Ref. Appendix G).

3.7 IOC Maintenance

IOC is considered the first iteration of the public-released CDR. As stated in the GCOS guidance, CDRs require periodic improvements to an ever expanding climate record. CDRs have to be updated as new observational datasets are acquired or new algorithms are developed. These updates may only be required to partly follow the R2O process. An example is a new dataset that needs to run on coded algorithms that remain unchanged. Regardless of the complexity of the CDR update, the key to maintaining transparency is a well documented audit trail.

3.8 IOC to FOC

The full progression from IOC to FOC will be detailed in a subsequent CDR guidelines document.

Appendices

Appendix A – CDR Assessment Example

Appendix B – Submission Agreement Example

Appendix C – Code Header Document Example

Appendix D – Data Flow Diagram Example

Appendix E – CDR Maturity Matrix

Appendix F – IOC Checklist

Appendix G – GCOS 12 Requirements for a CDR

Appendix H – Frequently Asked Questions

Appendix A – CDR Assessment Example

Passive MW Sea Ice: Assessment of CDR Readiness

by NCDC Scientific Point of Contact: Peggy Smith

Purpose

To try and gauge the level of effort needed to meet all the Initial Operating Capability (IOC) requirements

Also, from the R2O guidelines:

“Prior to committing significant internal resources to any main effort, a comprehensive CDR assessment must be conducted by the POC to determine whether the CDR is mature enough for transition.”

IOC Requirements

The following are the IOC requirements laid out by the CDR project:

- Code
  - Enter code in subversion
  - Document header information
  - Ensure code meets minimal standards outlined in guidance
  - Create README (cookbook) – Step by step instructions to run
  - Archive source code and README instructions
  - Make source code and README package available (web)
- Documents
  - Flow chart of process
  - Maturity matrix – level 3
  - Source code headers (robodoc)
  - CATBD written
  - Archive document package
  - Make documents available (web)
- Data
  - Submission agreement in place
  - Product in NetCDF format
  - Product meets metadata standards
  - Archive available Input/Ancillary data
  - Product archived
  - Product available (THREDDS)

Assessment

This assessment follows the structure of the IOC requirements.

Code

The code is written in different languages, mainly C and IDL; some PERL and shell scripts. The programming standards (headers and comments) are in place for the programs developed or improved at NSIDC. The same is expected but not sure to what degree for the part of the code supplied by people at the NASA Goddard (FORTRAN may be used for this portion of the code).

PI does not believe that it will be an issue for code-sharing, although he needs to get permission for the code supplied by people at the NASA Goddard. One possible option is for NSIDC to provide the access and support for the source code with the CDR program's approval.

Strengths

The NASA Team (NT) algorithm and the Bootstrap (BT) algorithm have been under development/application since 1980s. Both have been tested and validated extensively. The algorithm sensitivity and sources of errors are well-documented in publications.

The strength of NT is that it is not sensitive to changes in surface temperature while BT is less sensitive to thin ice and layering **within snow and ice.**

Weaknesses

NT is sensitive to thin ice and layering within snow and ice while BT is sensitive to changes in surface temperature.

Reliance on operational DMSP data is risky as errors do not get caught, corrected, or reported quickly and there is no reprocessing of bad data.

Pending issue

Need to get permission from people at the NASA Goddard for code-sharing for a portion of the code.

Documents

Maturity

The algorithms/data are fairly mature (level 3 or higher). The following table summarizes the maturity of the product by category:

Category	Rating	Justification
Sensor Use	**4**	Operational Missions: Nimbus-7 and DMSP; Sensors: SMMR: 1978 - 1987; SSM/I: 1987 - 2009; SSMIS: 2002 - ; AMSR-E: 2002 -
Algorithm Stability	**4**	However, this does not consider that the entire NCDC processing side (satellite pre and post processing) is not developed; Pending issue in code transferring.
Metadata	**3**	Research grade (extensive); Meets international standards
Documentation	**5**	Public ATBD, Operational Algorithm Description (OAD) and Validation Plan; Peer-reviewed algorithm, product and validation articles
Validation	**4**	Uncertainty estimated over widely distributed times/location by multiple investigators; Differences understood.
Public Release	**5**	Multi-mission record is publicly available with associated uncertainty estimate
Science and Applications	**6**	Used in various published applications and assessments by different investigators
Maturity	**3**	Lowest rating among categories

Data

The Sea Ice product has been routinely generated and distributed publicly at NSIDC. Initially, the products will be generated and hosted at NSIDC while the data will be ingested and stored on tapes at NCDC with a link pointing to NSIDC data server on the NCDC CDR website.

Risks

2007 QC-ed F-17 SSMIS input brightness temperature data not available, which is needed for sea ice inter-calibration using the Bootstrap (BT) algorithm.

Conclusion

I believe this product can be transitioned to a CDR within the required timeline. The algorithms are well documented, validated, and used by the community. The metadata associated with the products are FGDC and ISO 19115 compliant. The exception here will be the fact that the data will be generated and served from NSIDC, at least initially.

Appendix B – Submission Agreement Example

Submission Agreement Between the NCDC Stewardship Branch And the NCDC Archive Branch For Upper Tropospheric Water Vapor (UTWV) Products From HIRS Channel 12 Brightness Temperatures Including the Fundamental Climate Data Record (FCDR)

September 13, 2010

Introduction

This document represents the agreement that the National Climatic Data Center (NCDC) Stewardship Branch (the "Provider") and the NCDC Archive Branch (the "Archive") have reached for submitting the Provider's data, the Upper Tropospheric Water Vapor (UTWV) Products from High-Resolution Infrared Radiation Sounder (HIRS) Channel 12 Brightness Temperatures, including the Fundamental Climate Data Record (FCDR), to the Archive for long-term preservation. It represents a joint effort between the Provider and the Archive to accurately document the agreement and the expectations between the two groups.

Data Description

Intersatellite calibration is carried out for the UTWV data from clear-sky HIRS channel 12 measurement. As the intersatellite biases are scene brightness temperature dependent, an algorithm is developed to account for the varying biases with respect to brightness temperature. The bias correction data are derived from overlaps of monthly means of each 10-degree latitude belt. For the colder temperature range, data from the simultaneous nadir overpass observations are incorporated. The HIRS measurements from the NOAA series of polar orbiting satellites are calibrated to a baseline satellite. The time series of the intersatellite calibrated HIRS UTWV data from 1979 to present is constructed and anomaly data are computed.

- *Final edits to description for Archive Metadata are TBD*

Items to Be Submitted

1. HIRS UTWV for gridded Monthly Means, Daily Means, and Swath data in netCDF by version/revision ("v01r00" will be the initial archived version)
2. HIRS UTWV input/ancillary data files by version/revision (intermediate HIRS L1B channel 12 brightness temperatures and other data are TBD)
3. HIRS UTWV source code by version/revision
4. HIRS UTWV documentation by version/revision

Contacts

Data Provider Contacts:

Name
RSAD Deputy Chief
NCDC / RSAD
+1 828-271-4328
email

Name
Physical Scientist
NCDC / RSAD / SB
+1 828-350-2005
email

Name
IT Specialist
NCDC / RSAD / SB
+1 828-257-3017
email

Archive Contacts:

Name
Archive Branch Chief
NCDC / RSAD / AB
+1 828-271-4445
email

Name
IT Specialist
NCDC / RSAD / AB
+1 828-271-4368
email

Transfer Interface

The Provider will FTP push files to a directory on an NCDC server as specified by the Archive. The Provider will be responsible for verifying the success of an FTP session.

File Naming and Size

NetCDF Data:
Data files have the following naming convention:
HIRS-CH12_<*TYPE*>_v<*NN*>r<*NN*>_<*SATID*>_<*YYYY*><*DDD*>.nc

Where:

_ = file field delimiter
HIRS-CH12 = static field that identifies the HIRS channel 12 brightness temperature (UTWV) product series
<*TYPE*> = five to six characters identifying the data type, with the valid domain:
“MONGRD” for Monthly Means on a 2.5x2.5 degree grid
“DAYGRD” for Daily Means on a 2.5x2.5 degree grid
“SWATH” for aggregated granules at original HIRS L1B swath resolution (20km pixel)
v<*NN*>r<*NN*> = two-digit version number and two-digit release number corresponding with the data production
<*SATID*> = three character identifier for the POES or MetOp satellite that carried the HIRS instrument (only present for the Swath data type), with the valid domain:
“T-N” for TIROS-N
“N06” for POES-6
“N07” for POES-7
“N08” for POES-8
“N09” for POES-9
“N10” for POES-10
“N11” for POES-11
“N12” for POES-12
“N14” for POES-14
“N15” for POES-15
“N16” for POES-16
“N17” for POES-17
“M02” for MetOp-A
<*YYYY*> = four-digit year, from “1978” to “2009”
<*DDD*> = three-digit day of year (“001” - “366”) (field is only present for Swath data type)
nc = file extension for netCDF

Example netCDF file names:
HIRS-CH12_MONGRD_v01r01_2002.nc
HIRS-CH12_DAYGRD_ v01r01_2002.nc
HIRS-CH12_SWATH_ v01r01_N17_2002365.nc

Data are stored using netCDF-3.6.1. The gridded Monthly Means are yearly files with an average size of 1MB. The gridded Daily Means are also yearly files with an average size of TBD MB. There are approximately 25,000 daily Swath files for the 30+ year record with an average size of 2MB.

Data Tar Files:
The Swath data files will be archived in compressed yearly tape archive (tar) files by satellite with the following tar file naming convention:
HIRS-CH12_ SWATH_v<*NN*>r<*NN*>_<*SATID*>_<*YYYY*>.tar.gz

Where:
_ = file field delimiter
HIRS-CH12_SWATH = static field that identifies the HIRS channel 12 brightness temperature (UTWV) Swath product
v<*NN*>r<*NN*> = two-digit version number and two-digit release number corresponding with the data production
<*SATID*> = three character identifier for the POES or MetOp satellite that carried the HIRS instrument (only present for the Swath data type)
<*YYYY*> = four-digit year, from “1978” to “2009”
tar = file extension for tar archive file
gz = GNU zip compression extension

Example tar file name:
HIRS-CH12_SWATH_ v01r01_N17_2002.tar.gz

There will be no more than 366 Swath netCDF files in a yearly Swath tar file per satellite. Monthly Means and Daily Means will be archived as individual netCDF files and will not be archived as tar files.

Source Code:
Source code tar files will have the following naming convention:
hirs_ch12_source_code_v<*NN*>r<*NN*>.tar.gz

_ = file field delimiter
hirs_ch12 = static field that identifies the HIRS channel 12 brightness temperature (UTWV) product
source_code = static field that identifies source code for HIRS UTWV
v<*NN*>r<*NN*> = two-digit version number and two-digit release number corresponding with the data production of the same version/release
tar = file extension for tar archive file
gz = GNU zip compression extension

Example source code tar file name:
hirs_ch12_source_code_v01r01.tar.gz

Input/Ancillary Data:
Input/ancillary data files will have the following naming convention:
Archiving of intermediate HIRS products is TBD

Documentation:
Documents will have the following naming convention:
HIRS_UTWV_<*DOC_TYPE*>_ v<*NN*>r<*NN*>.<*ext*>

Where:
HIRS_UTWV = static field that identifies the HIRS UTWV product
<*DOC_TYPE*> = documentation type, with the domain:
“OAD” for Operational Algorithm Description
“Flowchart” for data flow processing diagram
“Code_Headers” for code headers (generated by ROBODoc software)
“Maturity_Matrix” for product’s maturity ratings
v<*NN*>r<*NN*> = two-digit version number and two-digit release number corresponding with the data production of the same version/release
<*ext*> = appropriate file extension for the document, e.g., “doc” or “pdf”

All archived documentation will be stored in a single tar file with the following naming convention:
HIRS_UTWV_DOC_v<*NN*>r<*NN*>.tar

Where:
HIRS_UTWV_DOC = static field that identifies documentation for the HIRS UTWV product
v<*NN*>r<*NN*> = two-digit version number and two-digit release number corresponding with the data production of the same version/release
tar = file extension for tar archive file

Example documentation tar file name:
HIRS_UTWV_DOC_v01r01.tar

Manifests:
A manifest file with file information and MD5 checksum value for each archived file is required for ingest of the product, input/ancillary data, and source code files.

The format of the information in a manifest file is:
file_name,file_size,MD5_checksum (three values comma delimited per row with no spaces)

A manifest can contain any number of file checksum values (i.e., one manifest for each file or one manifest for the entire record).

Manifests will have the following naming convention:
manifest_<*yyyymmdd*>_<*HHMMSS*>.<*ext*>

Where:
_ = file field delimiter
manifest = static field that identifies manifest file
<*yyyymmdd*> = manifest file creation date stamp for unique manifest file name
<*HHMMSS*> = manifest file creation time stamp for unique manifest file name
<*ext*> = appropriate file extension for manifest file, e.g., "txt" or "xml"

Example manifest file name:
manifest_20100810_101500.txt

A manifest file will be provided for each Monthly Mean and Daily Mean netCDF file as well as each Swath tar file. A manifest is also expected for any input/ancillary data, source code and documentation tar files.

Submission Schedule

Data submission for the initial version (i.e., v01r01) of the product and supporting information will begin no later than September 16, 2010 and will finish no later than September 24, 2010. Source data and code for the corresponding initial version will also be submitted during September 2010.

Additional product versions are expected to be delivered as requested by the Provider and these will be archived in a similar fashion as the initial submission. A routine schedule may be established once the product is fully operational.

Validation

The Archive will use file name, size and MD5 checksum to verify the integrity of a delivered file.

Error Reconciliation

The Archive will report any unexpected file size or a duplicate file for a version to the Provider. The same procedure will be true for any file integrity or checksum error. A new corresponding manifest file will be required for re-submitted files from the Provider.

Confirmation

The Archive will provide confirmation of successful data ingest for a version or as requested by the Provider.

Quality Assurance

No Quality Assurance will be performed by the Archive.

Archive Storage

All files will be stored on the NCDC HDSS under DSI:
3629_<*XX*>
Where, <*XX*> is the two-digit DSI subset number.

The subset number will increment by one (e.g., “01”, “02” … “99”) with each new product release, however, a subset number will not necessarily correspond to the product version or release number stored under that subset.

Directory structure for the data will be by year and contain the Monthly Means and Daily Means netCDF files and the compressed yearly Swath tar files (all three data types):
/aab/36xx/3629_<*XX*>/CDR/<*YYYY*>/HIRS-CH12_MONGRD_v<*NN*>r<*NN*>_<*YYYY*>.nc
/aab/36xx/3629_<*XX*>/CDR/<*YYYY*>/HIRS-CH12_DAYGRD_v<*NN*>r<*NN*>_<*YYYY*>.nc
/aab/36xx/3629_<*XX*>/CDR/<*YYYY*>/HIRS-CH12_SWATH_ v<*NN*>r<*NN*>_T/N/M<??>_<*YYYY*>.tar.gz

Directory structure for ancillary or input data, source code, and documentation, respectively:
/aab/36xx/3629_<*XX*>/ANC/TBD
/aab/36xx/3629_<*XX*>/SRC/hirs_ch12_source_code_v<*NN*>r<*NN*>.tar.gz
/aab/36xx/3629_<*XX*>/DOC/HIRS_UTWV_DOC_v<*NN*>r<*NN*>.tar

Note that ancillary data archived under existing DSI directories will only be referenced.

Constraints

No constraints on data access, use or other.

Access and User Services

The Archive will write metadata that follows the FGDC CSDGM content standard with Extensions for Remote Sensing using information from the Provider. This metadata file, with the ID TBD, will be stored and published in the NOAA Metadata Manager Repository (NMMR) and from there harvested to metadata portals and clearinghouses for public users (URL: http://www.ngdc.noaa.gov/metadata/published/NCDC/TBD). User community access will be through THREDDS for data on ‘eclipse’, and access to data on tape will be through the HDSS Access System (HAS) - TBD. Provision of archived data on media will be provided through NCDC Customer Services.

Additional Terms

None

Appendix C – Code Header Document Example

TABLE OF CONTENTS

1. PROJECT/ERSST
1.1. ERSST/dati2.upd.situ.v3b.f
.
.
.
1.31. ERSST/sst2d.situ.v3b.f/intts
1.32. ERSST/ssta.merg.situ.v3b.f/range

1. PROJECT/ERSST [PROJECTS]

[Top] [PROJECTS]

NAME

ersstv3b_oper_situ_only_col.sh

PURPOSE

To generate the analyzed Extended Reconstruction Sea Surface Temperature (**ERSST**) on a 2-deg grid from in situ data (ship and buoy, NOT satellite: v3b uses in situ only),and transfer to distribution directories.

DESCRIPTION

This is the main script that launches a series of fortran programs for computing for a specific month and year (determined from the current machine date). The operational runs will affect values in recent past years due the long-term averaging

Also, the program uses output from a one-time climatological run (1880 to around 1985). Most programs write output for all the years sequentially in one binary file. However, depending on the program, the month processed may just be added to the pre-existing file, or the entire file may be rewritten from 1985 onward.

The processing is as follows:
First, in situ data (ship and buoy) is ftp'd from source locations. Adjustments are made for distance of point obs from grid center, difference in dependability of ship and buoy data, etc. and other quality checks are made. The data is placed on a 2-deg grid, and anomalies are computed. Adjustments are also made for sea ice presence. The sea ice data comes from the daily OISST analysis. Statistical analysis is done in 2 steps:
1)The decadal or low frequency component is determined from the anomalies and then the residuals are computed
2)The high frequency analysis is performed on the residuals.
The **ERSST** is then computed from the sum of the two components and error variance is estimated. After the **ERSST** computation, other programs are run to update related products(land and merged land-ocean SST) that use **ERSST**.

These other products are continually updated on a different schedule, external to this script. Areal averages are computed for the Climate Monitoring group, and plots are made to check the **ERSST** output. Comparisons with v2 and the v3 (satellite) are also made, that are produced separately.

AUTHOR

Chunying Liu

CREATION DATE

04/01/2008

COPYRIGHT

THIS SOFTWARE AND ITS DOCUMENTATION ARE CONSIDERED TO BE IN THE PUBLIC DOMAIN AND THUS ARE AVAILABLE FOR UNRESTRICTED PUBLIC USE. THEY ARE FURNISHED "AS IS." THE AUTHORS, THE UNITED STATES GOVERNMENT, ITS INSTRUMENTALITIES, OFFICERS, EMPLOYEES, AND AGENTS MAKE NO WARRANTY, EXPRESS OR IMPLIED, AS TO THE USEFULNESS OF THE SOFTWARE AND DOCUMENTATION FOR ANY PURPOSE. THEY ASSUME NO RESPONSIBILITY (1) FOR THE USE OF THE SOFTWARE AND DOCUMENTATION; OR (2) TO PROVIDE TECHNICAL SUPPORT TO USERS.

INPUTS

buoyship_quarter/nqyyyymm.Z - ftp ship and buoy data
Outputs from the Fortran programs are passed to other Fortran programs with each program building/modifying the data for the next program

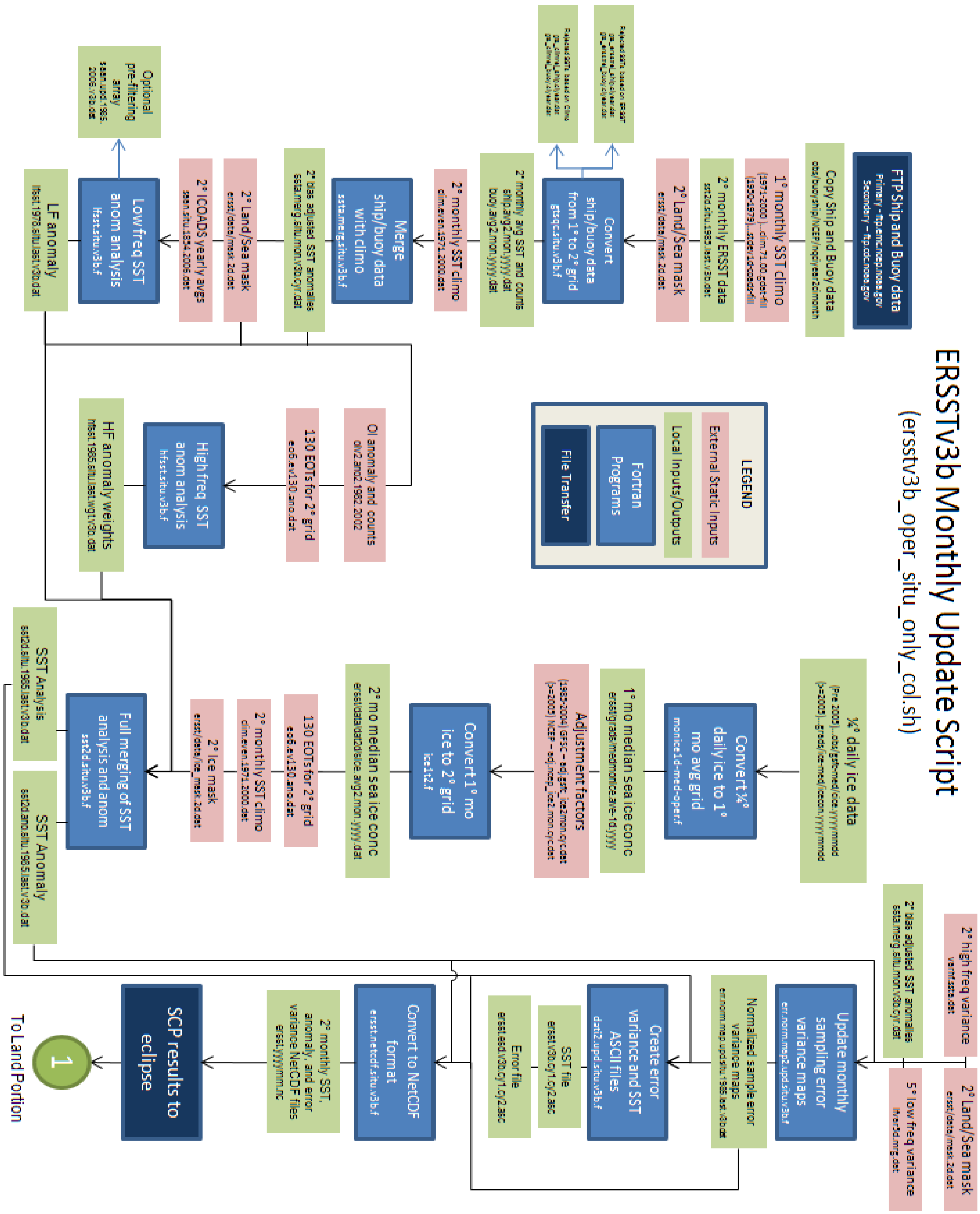
OUTPUTS

Updated **ERSST** integer and data files **ERSST**/datat/**ERSST**-v3b/situ/**ERSST**.v3b.yr1.yr2.asc **ERSST**/datat/**ERSST**-v3b/situ/**ERSST**.esd.v3b.yr1.yr2.asc
NetCDF formatted **ERSST** file **ERSST**/data/netcdf-v3b/situ/**ERSST**.yyyymm.nc

MODIFICATION HISTORY

04/30/2008 – C. Liu modified from /raid2/**ERSST**/ftn/ersstv3b_oper_situ_only_col.sh to remove satellite data
08/28/2009 – V. Banzon added comments
08/17/2010 – V. Banzon removed Land and merged comments after C. Liu removed related codes The conditions for writing ascii for previous decade were put in but left commented since that was not in the original script

Appendix D – Data Flow Diagram Example



Appendix E – CDR Maturity Matrix

Maturity	Sensor Use	Code Stability	Metadata & QA	Documentation	Product Validation	Public Access	Applications
1	Conceptual development	Little or none	Draft Climate Algorithm Theoretical Basis Document (C-ATBD); paper on algorithm submitted	Little or None	Restricted to a select few	Little or none	Conceptual development
2	Significant code changes expected	Research grade	C-ATBD Version 1+ ; paper on algorithm reviewed	Minimal	Limited data availability to develop familiarity	Limited or ongoing	Significant code changes expected
3	Moderate code changes expected	Research grade; Meets int'l standards: ISO or FGDC for collection; netCDF for file	Public C-ATBD; Peer-reviewed publication on algorithm	Uncertainty estimated for select locations/times	Data and source code archived and available; caveats required for use.	Assessments have demonstrated positive value.	Moderate code changes expected
4	Some code changes expected	Exists at file and collection level. Stable. Allows provenance tracking and reproducibility of dataset. Meets international standards for dataset	Public C-ATBD; Draft Operational Algorithm Description (OAD); Peer-reviewed publication on algorithm; paper on product submitted	Uncertainty estimated over widely distributed times/location by multiple investigators; Differences understood.	Data and source code archived and publicly available; uncertainty estimates provided; Known issues public	May be used in applications; assessments demonstrating positive value.	Some code changes expected
5	Minimal code changes expected; Stable, portable and reproducible	Complete at file and collection level. Stable. Allows provenance tracking and reproducibility of dataset. Meets international standards for dataset	Public C-ATBD, Review version of OAD, Peer-reviewed publications on algorithm and product	Consistent uncertainties estimated over most environmental conditions by multiple investigators	Record is archived and publicly available with associated uncertainty estimate; Known issues public. Periodically updated	May be used in applications by other investigators; assessments demonstrating positive value	Minimal code changes expected; Stable, portable and reproducible
6	No code changes expected; Stable and reproducible; portable and operationally efficient	Updated and complete at file and collection level. Stable. Allows provenance tracking and reproducibility of dataset. Meets current international standards for dataset	Public C-ATBD and OAD; Multiple peer-reviewed publications on algortihm and product	Observation strategy designed to reveal systematic errors through independent cross-checks, open inspection, and continuous interrogation; quantified errors	Record is publicly available from Long-Term archive; Regularly updated	Used in published applications; may be used by industry; assessments demonstrating positive value	No code changes expected; Stable and reproducible; portable and operationally efficient

1 & 2	**Research**
3 & 4	**IOC**
5 & 6	**FOC**

Appendix F – IOC Checklist

IOC Requirements (v 1.0)	Primary Responsibility	CDR Name v01r00
Code		
Document Header information (comments) in source code	PI	
Deliver source code*	PI, POC	
Enter source code into subversion for security review	OPS	
Conduct security review	IT	
Create README (Cookbook) - Step by step run instructions	PI	
Deliver README*	PI, POC	
Archive source code and README package (part of SA)	Archive, OPS, POC	
Make source code and README package available (web)	Access, Archive	
Documents		
Create Flow Chart of process	PI	
Fill out Maturity Matrix (MM)- lvl 3 min	PI	
Create ATBD	PI	
Deliver Flow Chart, MM, and ATBD*	PI, POC	
Create source code header document (robodoc)	OPS	
Archive Document Package (4 docs archived together)	Archive, OPS, POC	
Make Documents available (web)	Access, Archive	
Data		
Convert to NetCDF	PI, OPS	
Create Submission Agreement (SA) - details all deliveries	Archive, PI, POC	
Create Metadata for CDR product	Archive, PI, POC	
Deliver CDR product* (and ancillary data if req)	PI, OPS	
Archive available Input/Ancillary Data (details in SA)	Archive, OPS	
Archive CDR Product (details in SA)	Archive, OPS	
Make CDR Product available (web)	Access, Archive	
	PI	
POC should facilitate at all steps of the R2O process	POC	
*Deliveries will be reviewed and accepted/archived or returned for revision/modification		

Appendix G – GCOS 12 Requirements for a CDR[1]

1. Full description of all steps taken in generation of FCDRs and ECV products, including algorithms used, specific FCDRs used and characteristics and outcomes of validation activities.
2. Application of appropriate calibration/validation activities.
3. Statement of expected accuracy, stability and resolution (time, space) of the product including a comparison with the GCOS requirements.
4. Assessment of long-term stability and homogeneity of the product.
5. Information on the scientific review process related to FCDR/product construction (including algorithm selection), FCDR/product quality and applications.
6. Global coverage of FCDRs and products where possible.
7. Version management of FCDRs and products where possible.
8. Arrangements for access to the FCDRs, products and all documentation.
9. Timeliness of data release to the user community to enable monitoring activities.
10. Facility for user feedback.
11. Application of a quantitative maturity matrix if possible.
12. Publication of a summary (webpage or peer-reviewed article) documenting point-by-point the extent to which this guideline has been followed.

Table 1. GCOS 12 Requirements for CDRs[1]

[1] Guidelines for the Generation of Datasets and Products Meeting GCOS Requirements, GCOS-143 (WMO/TD No. 1530), May 2010

Appendix H – Frequently Asked Questions

Q1. What is the role and structure of the CDR Program Office?

A1. RSAD Management has implemented a matrix management approach to CDR acquisition and implementation. In short, that means RSAD branches and offices all contribute personnel to task-specific teams, since no single branch or office has all of the needed personnel or expertise to be successful. The CDR Program Office serves a CDR planning and management function for RSAD and NCDC. As such, it typically leads RSAD's CDR-related teams. It is also responsible for the development of definitions, requirements, guidelines, procedures and processes – including system architecture analysis and definition -- and manages the implementation and acquisitions as appropriate. This scope includes the business affairs such as performance metrics, budgets, grants and contracts management.

Q2. When there is an interim delivery, is there a defined space to put code, netCDF example, etc. This implies it will sit there to be examined by Archive, Operations, Products Branch, and/or IT security?

A2. A dedicated space on the ftp server will be established so PIs or POCs can stage code, documents, and data for review and verification. The proper place for source code needing a security review is in subversion on the *conman* server.

Q3. How is information communicated up the CDR management chain from the POC's?

A3. POCs provide weekly updates to the CDRP R2O project manager on their specific CDR's status. The R2O project manager carries unresolved questions directly to the standing CDRP/Branch Manager meeting and informs the POCs of any solutions or future actions discussed in the meeting. Quarterly reviews are also held with all POCs and managers present for discussions of issues/concerns and to provide decisions or guidance.

Q4. Is there one location to get all the CDR guidelines and standards?

A4. The latest versions of all guidelines, standards, and templates are available on the CDRP web site, under the *'Development Guidelines'* (http://www.ncdc.noaa.gov/cdr/guidelines.html).