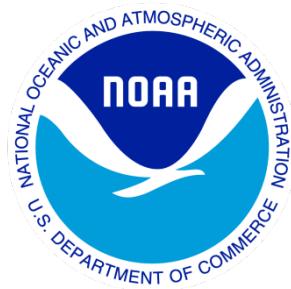

Climate Data Record (CDR) Program

IOC to FOC Transition Process



CDR Program Document Number: CDRP-PRCS-0001
Originator Document Number: TBD
Version 1.0 / July 5, 2011

TABLE of CONTENTS

1. INTRODUCTION	4
1.1 Executive Summary	4
1.2 Goal.....	5
1.3 Scope and Depth	5
2. CDR AT INITIAL OPERATING CAPABILITY.....	5
2.1 Background	5
2.2 Description of IOC	6
2.2.1 Assessment.....	6
2.2.2 Preparation.....	6
2.2.3 Transfer	6
2.2.4 Verification.....	6
2.2.5 Archival.....	6
2.2.6 Access.....	7
3. CDR AT FULL OPERATING CAPABILITY	7
3.1 Background	7
3.2 Description of FOC.....	7
3.2.1 Reproducibility.....	7
3.2.2 Verification.....	7
3.2.3 Extensibility	8
3.2.4 Preservation	8
3.2.5 Accessibility.....	8
4. DESCRIPTION OF THE IOC TO FOC TRANSITION PROCESS	8
4.1 Background, Objectives, and Scope	8
4.2 FOC Phases.....	9
4.2.1 Transition.....	9
4.2.2 Production	9
4.3 Users	9
4.4 Roles and Responsibilities for Transition from IOC to FOC	9
5. OPERATIONAL SCENARIOS	11
5.1 Transition Phase	11
5.1.1 Assessment.....	12
5.1.2 Planning	12
5.1.3 Preliminary test.....	12
5.1.4 Development	13
5.1.5 First level system integration and test.....	13
5.1.6 Verification and validation	14
5.1.7 Second level system and integration test	14
5.2 Production Phase	15
5.2.1 Planning	15
5.2.2 Production	15
5.2.3 Operational monitoring.....	15
5.2.4 Preservation	15

5.2.5	Accessibility.....	16
5.2.6	Re-evaluation	16
6.	IOC TO FOC PROCESS FLOW.....	17

LIST of FIGURES

Figure 1:	IOC to FOC Process Flow.....	17
------------------	-------------------------------------	-----------

ACRONYMS and ABBREVIATIONS

Acronym or Abbreviation	Meaning
C-ATBD	Climate-Algorithm Theoretical Basis Document
CDR	Climate Data Record
CIR	Climate Information Record
CM	Configuration Management
CSMD	Climate Services and Monitoring Division
DAAB	Data Access and Applications Branch
ECV	Essential Climate Variable
FCDR	Fundamental Climate Data Record
FOC	Full Operating Capability
IOC	Initial Operating Capability
IT	Information Technology
NCDC	National Climatic Data Center
NetCDF	Network Common Data Form
NMMR	NOAA Metadata Manager Repository
NOAA	National Oceanic and Atmospheric Administration
OAD	Operational Algorithm Description
PI	Principal Investigator
POC	Point of Contact
RSAD	Remote Sensing and Applications Division
R2O	Research to Operations
SIT	System Integration and Test
TCDR	Thematic Climate Data Record

1. Introduction

1.1 Executive Summary

The National Climatic Data Center (NCDC) has been charged with creating, maintaining, and securing scientifically defensible multi-decadal climate data records (CDRs) which may be used to inform high-level policy decisions. Currently, a process exists to bring research quality CDRs to initial operating capability (IOC) status (CDRP-PLAN-0017). At IOC, a CDR, including its data, source code, and documentation, is archived and is available to the public, but its data may or may not have been independently reproduced. A CDR at IOC is not regularly maintained nor is it necessarily extended as new data become available. Moving CDRs currently at IOC status to full operating capability (FOC) is the next step in the evolution of a capability to provide uninterrupted and continuous climate data records.

The CDR Program is responsible for a process to transition CDRs at IOC to FOC. At FOC, a CDR will be systematically and operationally maintained in a cycle of continuous improvement. The transition process is envisioned as being progressive, with the CDR meeting increasingly rigorous standards until FOC is attained. At any time, the maturity of a CDR can be evaluated against the CDR Maturity Matrix (CDRP-MTX-0008), with the CDR Maturity Evaluation Criteria (CDRP-GUID-0020) providing quantifiable measures of progress.

FOC is envisioned as having two phases: the first, a transition phase to ensure that source code and documentation meet CDR Program standards and that results are reproducible; and the second, a production phase to systematically and routinely extend the record as new data become available or to reprocess the record if recalibrated data become available. The process also includes the potential for a CDR to re-enter the transition phase if a significantly improved algorithm becomes available.

Both Fundamental CDRs (FCDRs), which contain data but are not associated with a particular Essential Climate Variable (ECV), and Thematic CDRs (TCDRs), which are related to ECVs, will be maintained and extended in as close to real time as possible. The same process will serve for both. (For definitions of these and other CDR related terms, refer to the CDR Program Glossary, CDRP-MISC-0045).

Number: CDRP-PRCS-0001

Originator: Rebecca Harris, GST/NCDC, 828-271-4905

Sponsor: CDR Program Office

1.2 Goal

The goal of this document is to clearly delineate the steps, procedures, and thresholds needed to bring a CDR to FOC. In addition, this document provides information that can be used by the CDR Program and NCDC to evaluate the cost and complexity of the IOC to FOC transition, to identify potential technology and other resourcing needs, and to enable planning and budgeting for production, maintenance, and distribution of future CDRs. It defines a process that will allow the National Oceanic and Atmospheric Administration (NOAA) to systematically and routinely produce scientifically defensible CDRs that are extensible, are securely archived, and are available to the public.

1.3 Scope and Depth

This document presents a view of the IOC to FOC transition process at a high level, with general descriptions of some subordinate processes. This document contains a description of IOC (the point at which CDRs will enter the transition process), a description of FOC (the desired outcome of the transition process), and a description of the process itself. Stakeholders, as well as their roles and responsibilities, are identified.

2. CDR at Initial Operating Capability

2.1 Background

The CDR Program of the Remote Sensing and Applications Division (RSAD) of NCDC is responsible for the development of CDRs. The CDR Program competitively selects Principal Investigators (PIs) to perform the research and development necessary to bring a scientifically defensible CDR to the IOC level. CDR development follows a consistent and well-defined set of improvement milestones that are described in the CDR Maturity Matrix and the CDR Maturity Evaluation Guidelines. The CDR Program also maintains standards for metadata and software coding that must be met by CDRs at IOC level. To achieve IOC, the archived CDR package must include a Climate Algorithm Theoretical Basis Document (C-ATBD) that describes the theoretical basis of peer-reviewed algorithm(s) that have been used to generate the output dataset. This algorithm will have been tested by the PI, who will also have confirmed that its results are stable. The source code, output data, and documentation will all have been archived at NCDC. Although the data may not have been independently reproduced, it should be possible, given the archived documentation, to reproduce the CDR output data using the same code on the same computing platform. Code condition may vary, but minimum CDR Program standards for IOC will have been met. Metadata will be in netCDF format and meet NCDC metadata standards.

2.2 Description of IOC

All CDRs at IOC will have completed the following steps: assessment, preparation, transfer, verification, archival, and access.

2.2.1 Assessment

Assessment is based upon a significant portion of the code, the draft C-ATBD, the data flow diagram for the software, and sample output data. The feasibility and complexity of the transition to IOC will have been considered, and the volume of input and output data will have been estimated. Source code may or may not meet minimum CDR Program standards as described in CDR General Software Coding Standards (CDRP-STD-0007). At IOC, moderate changes to source code can still be expected before it can become fully operational.

2.2.2 Preparation

At IOC, a CDR package, including input, output, and ancillary data, source code, and documentation, will have been prepared. Required documents include a C-ATBD, source code headers, a README file with step-by-step run instructions, a data flow diagram, and a completed CDR Maturity Matrix. Metadata will exist at the file and collection level and are sufficient for use and understanding of the data independent of external assistance. Metadata may not yet be complete. All file names will follow NCDC Archive conventions, and CDR data will be in netCDF format. Source code is expected to be documented, although detail and extent of the documentation may vary.

2.2.3 Transfer

At IOC, the CDR package will have been transferred by the PI to NCDC per the Submission Agreement. Source code and documentation will have been placed under version control.

2.2.4 Verification

The transfer of the CDR package will have been verified via a combination of MD5 checksum, file name, and file size, all of which will have been provided by the PI to the Archive in a manifest file. A basic security review, for informational purposes only, will have been performed on the source code, but the source code will not have been compiled or executed at NCDC and so is not required to have passed the security review. Data are affirmed by the PI as reproducible when using the same inputs, same source code, and same platform. Data may not have been independently reproduced by a third party.

2.2.5 Archival

The CDR package will have been archived at NCDC in accordance with the Submission Agreement.

2.2.6 Access

The CDR package will have been made available via NCDC's online data access tools or via the CDR Program's website (<http://www.ncdc.noaa.gov/cdr/index.html>). There are no constraints on access or use. A discovery metadata file will have been published on the NOAA Metadata Manager Repository (NMMR).

3. CDR at Full Operating Capability

3.1 Background

NCDC will be responsible for routinely and systematically generating CDR data using codes and systems that conform to the CDR Programs standards. The actual production of CDR data may take place within NCDC or at another suitable facility. To achieve FOC status, the CDR must have achieved maturity level 5 in all categories in the CDR Maturity Matrix and meet all level 5 CDR Maturity Evaluation Guidelines. At FOC, the reproducibility of the CDR data will have been verified. At FOC, any uncertainties will have been estimated over most conditions by multiple investigators and documented in peer-reviewed literature. At FOC, code will also have passed NCDC security review.

3.2 Description of FOC

To achieve FOC status, a CDR must have the following attributes:

3.2.1 Reproducibility

At FOC, the CDR's source code will have passed any required security review and will have been successfully made operational, with any platform dependencies identified. The data generated will be stable and within machine rounding errors of pre-determined test data sets for the CDR.

3.2.2 Verification

At FOC, source code, documentation, and metadata will meet CDR Program standards for FOC and will be scientifically defensible. Any discrepancies from original results will have been quantified and explained. Operational monitoring of CDR data will be in place. CDR data will be periodically reviewed against pre-determined test data sets and may be reprocessed when improved algorithms or newly recalibrated data become available.

3.2.3 Extensibility

At FOC, a CDR spans multiple decades and is capable of being extended in time to generate a stable climatic time series. CDRs will be routinely extended as new data become available from existing or new sensors, providing long-term product consistency. Forward processing of data will be performed in as near to real-time as possible, allowing time for assessment and correction of any problems with the data. As new algorithms become available, new TCDRs may be created with different theoretical bases or the number of CDRs may be increased to encompass additional ECVs.

3.2.4 Preservation

The CDR package will be preserved in accordance with NCDC standards and includes data (output data, ancillary data, and any input data not currently archived); metadata; production software source code; software to read the data; documentation; calibration/validation information and data; and quality assurance information. The C-ATBD and Operational Algorithm Description (OAD) of the CDR are preserved to provide more complete provenance traceability of the scientific basis of the CDR. The CDR will meet NCDC Archive standards for both data and media format.

TCDRs that are determined to no longer provide utility may be moved to passive storage or decommissioned.

3.2.5 Accessibility

Access to and distribution of CDRs to the user community will be assured and facilitated. CDRs will be available in as timely a manner as possible with user requests typically fulfilled within 24 hours. Access will be full and open with a provision for users to provide feedback. Data will be available in widely utilized formats, and proprietary environments for data usage will be avoided. Appropriate data manipulation and visualization tools may be provided.

4. Description of the IOC to FOC Transition Process

4.1 Background, Objectives, and Scope

This section of this document describes a process through which CDRs at IOC status will complete a transition to FOC. At FOC, a CDR will be systematically and routinely generated using codes and systems that conform to the CDR Program's coding and documentation standards. This process will help meet the United States government's goal of acquiring and archiving uninterrupted and indefinite climate data records that have been rigorously validated and are available to the public for utilization in further research or in decision support systems.

4.2 FOC Phases

There are two FOC phases, transition and production.

4.2.1 Transition

During the transition phase from IOC to FOC, CDR source code will be brought up to CDR Program standards for FOC and transformed into operational code that is both stable and portable. Steps will be taken to ensure that original results are reproducible and that life cycle costs can be minimized. The transition phase will take place primarily in a development environment and will occur on the platform most suitable for effecting the advancement of the CDR to FOC.

4.2.2 Production

Once FOC is achieved, the CDR will enter the production phase. Production processing will take place in an operational environment and will occur on the platform most suitable for secure and efficient production and maintenance of the CDR. Production runs will be of two general types: forward processing, which extends an existing CDR data record as new data become available, and re-processing, which creates a new version of a CDR when an algorithm is improved or when an algorithm with a different theoretical basis becomes available. Forward processing may occur on a regularly scheduled basis; re-processing will be done on a pre-approved basis.

Forward processing will include some capability for operational monitoring of output data, so that data anomalies may be recognized and examined. Forward processing is done in as near to real time as possible, allowing sufficient time for any required pre-processing of input data.

Re-processing of a CDR's input data may be desirable as scientific insights change over time and algorithms improve. A CDR whose algorithm has a new theoretical basis must complete the transition phase before entering production. Depending on the circumstance, the pre-existing CDR may also continue to be extended with the pre-existing algorithm.

4.3 Users

Classes of users that will request and utilize CDRs include, but are not limited to, the following: science, industry, business, government, and interested members of the public.

4.4 Roles and Responsibilities for Transition from IOC to FOC

The CDR Program, which is organized as an element of the NCDC Remote Sensing and Application Division (RSAD), is responsible for the transition of select CDRs to FOC

status. The development of CDRs necessarily requires the commitment of other RSAD branches as well as other NCDC divisions. Continuing support from the scientific community will also be needed.

The CDR Program and the Operations Branch of RSAD will determine whether the transition of a CDR should occur on site at NCDC or whether any part would better be accomplished at an appropriate offsite location. In either case, the CDR Program will oversee and track progress of the transitioning CDR and will lead readiness reviews. Once the transition to FOC has been accomplished, the CDR Program, with the help of the Operations Branch, will determine the appropriate location for production of the CDR. If an offsite environment is selected for either transition or production phase, the CDR Program will be responsible for coordinating the development of contracts with required resources. In conjunction with the scientific community, the CDR Program will determine the criteria and schedule for extending or reprocessing CDRs. In order to uphold CDR Program transparency standards, CDR Program procedures will be followed to document all decisions made. The CDR Program will work with an appropriate source to design and develop visualization tools that will make CDR data more accessible and useful to its user community. The CDR Program will also gather metrics on CDR usage and will establish mechanisms by which the user community can provide feedback on the CDR products.

The Operations Branch will assess the CDR package, including degree of compliance of source code with CDR Program software coding standards, and will estimate effort required to bring that code from IOC to FOC standards. The Operations Branch may also advise on whether any part of the transition or production effort is better performed outside NCDC, and, if the transition or production occurs within NCDC, the Operations Branch will support the effort to make CDR source code fully operational. The Operations Branch will be responsible for creating the Operational Algorithm Description (OAD) and will participate in readiness reviews led by the CDR Program.

The Information Technology (IT) Branch of NCDC's Support Services Division will facilitate and provide infrastructure (including version control) support for any transition or production of a CDR that occurs on site at NCDC. The IT Branch's Configuration Management (CM) team will provide assistance to the CDR Program's CM team to ensure their process meets NCDC's CM standards. The IT Branch's Security team will be responsible for computer security for the source code of any new CDR whose development or production occurs in-house at NCDC. The IT Branch will run security reviews on source code as needed. IT Branch will also participate in readiness reviews.

The Products Branch of RSAD and the scientific community will be engaged by the CDR Program to ensure that the CDR will meet, or continue to meet, CDR Maturity Evaluation Guidelines for FOC. They will assist with selection of test data sets, to enhance file level metadata or CDR documentation to the CDR Program's standards for FOC, to provide independent validation, to quantify biases and errors, and to advise on development of a method of operational monitoring appropriate to the CDR. The Products Branch will assist the Operations Branch in writing the OAD.

The Archive Branch of RSAD is responsible for preservation of any interim versions of the CDR as it moves through the transition to FOC as well as of all production versions of the CDR. Archive Branch will provide support in determining what metadata are needed at the file and collection level to continue to facilitate use and discovery by others or to allow provenance tracking.

The Data Access and Applications Branch (DAAB) of CSMD is responsible for providing public access to any versions of the CDR that are archived. DAAB will assist the CDR Program in monitoring data usage of TCDRs, in quantifying citations of CDR algorithms and products, and in developing and maintaining a means of eliciting feedback on CDRs from the scientific community.

The scientific community will be engaged by the CDR Program to assist in both transition and production phases. During the transition phase, the scientific community will advise in selection of “golden” data sets for testing, determination of criteria for validation, quantification of errors and biases, and decisions on operational monitoring. During the production phase, the scientific and user communities will assist in determining the criteria and schedule for extending or reprocessing CDRs as well as what portion of the CDR should be reprocessed. When new satellite data become available, the scientific community will assist the CDR Program in determining whether the current algorithm should be used to extend the CDR or whether data should be kept separate pending development of a new algorithm. The scientific community will create Climate Information Records (CIRs), which provide information about environmental phenomena of importance to science and society. The scientific and user communities will also be relied upon to monitor satellite and sensor characteristics and assist with calibration or recalibration of sensor data that may impact the validity of CDRs.

5. Operational Scenarios

5.1 Transition Phase

The CDR transition phase encompasses all efforts required to bring the CDR from IOC to FOC. It is assumed that every CDR entering the transition phase meets minimum CDR Program Maturity Evaluation Guidelines for IOC. Entry to the transition phase may occur under one of three circumstances:

- First entry by a new CDR which has completed the formal research to IOC process described in section 2.2 above. This includes a new CDR for the same variable as an existing CDR but with a new algorithm theoretical basis.
- Re-entry by a CDR with a revised algorithm having the same theoretical basis. Development of the revised algorithm may have been triggered by a need to improve the existing CDR or by a need to incorporate new satellite data into the existing CDR. In either case, all or some of the existing CDR will need to be

reprocessed with the new algorithm. This CDR may not have been through the formal research to IOC process but will meet at least minimum CDR Program Maturity Evaluation Guidelines for IOC. The reprocessed CDR will supersede the pre-existing CDR, although the previous version will be retained in the archive.

The following steps comprise the transition process:

5.1.1 Assessment

Because CDRs will enter the transition phase at varying levels of maturity, each CDR must be individually assessed:

- a. Confirm the completeness of the CDR package stored in the NCDC Archive.
- b. Evaluate the maturity of source code and documentation through automated tools and/or by visual inspection.
- c. Review the security check performed at IOC or perform an additional security check on source code, noting any problems.
- d. Determine the size of the source code.
- e. Determine the complexity of compute resources utilized in creation of the CDR by evaluating the previous development environment, size of input and output files, degree of parallelization of code, and compute time required.
- f. Verify the period of record and maturity of the sensors that were the source of input data and assure that all input data are available.
- g. Determine any special requirements for commercial software packages, interfaces, third-party libraries, etc. that deviate from CDR Program standards.
- h. If the new algorithm is for re-processing of an existing CDR, determine what portion of the CDR will be re-processed.

5.1.2 Planning

Once assessment is complete, plan the work required to achieve FOC.

5.1.3 Preliminary test

As necessary, confirm that data are reproducible. A complete CDR package will include a README file that contains step-by-step instructions for reproducing the output data.

- a. In the selected development environment, compile source code and bring it to a runnable state with minimal changes.
- b. Perform test run and confirm results with appropriate test data set.
- c. Update documentation to reflect any changes made. Log changes to source code.
- d. Utilize version control for all source code changes and new data products.
- e. Revise estimate of effort and schedule if necessary.

5.1.4 Development

Development is an iterative process and will be complete only when the CDR package has been thoroughly reviewed and determined to be ready for the next step. Source code may be subdivided into smaller units or subsystems for development and unit test.

- a. Modify or rewrite source code to meet CDR Program standards, guidelines, and recommendations. Modify or rewrite source code to meet required security standards.
- b. Add quality flags as appropriate.
- c. Modify or rewrite documentation to meet CDR Program standards.
- d. Update documentation to reflect any changes made. Log all changes to source code.
- e. Update metadata or add fields so that metadata are sufficient for provenance tracking and/or data use and discovery by others.
- f. Utilize version control for all source code changes and new data products.
- g. Unit test throughout and evaluate results.
- h. Record any anomalies, corrective actions, or lessons learned. Update test scripts as appropriate.
- i. Conduct final readiness review.

5.1.5 First level system integration and test

First level system integration and testing should be performed in an environment as close to the anticipated production environment as possible. It may or may not be an iterative process, depending upon the complexity and maturity of the source code. If any deviation from the original IOC results is found, the process returns to the development step for resolution before proceeding.

- a. Repeat security check on source code. If any significant problems are found, the process returns to the development step for resolution before proceeding.
- b. Determine the integration sequence. Assemble and integrate as necessary any smaller code units or subsystems into increasingly larger or more complex subsystems.
- c. In the selected environment, perform tests and capture results. Evaluate output against “golden” data sets and evaluation criteria defined in test plans. If any significant deviation from “golden” data sets or if evaluation criteria are not met, the process returns to the development step for resolution before proceeding.
- d. Measure performance and gather performance statistics. If any performance problems are found, the process returns to the development step for resolution before proceeding.

5.1.6 Verification and validation

Results will be verified against CDR Program requirements and validated against IOC results and/or “golden” data sets. Any problems found are recorded as defects, and the process returns to the development step.

- a. Verify that source code, documentation, and metadata are complete, correctly versioned, and meet required CDR Program standards. Verify that changes from the original IOC CDR package are well documented in accordance with CDR Program requirements.
- b. Validate output against expectations. Output data that meet criteria pre-determined by the scientific community will constitute acceptance testing.

5.1.7 Second level system and integration test

Source code must be tested and validated on the production platform selected before the CDR moves into production. The selection or creation of the production environment should occur well ahead of this step.

- a. Repeat all four steps of the system test in production environment.
- b. Repeat verification and validation steps. Output data that meet criteria pre-established by the scientific community will constitute acceptance testing.
- c. Formulate a submission agreement between the CDR Program and the NCDC Archive for new CDR package. Transfer the new CDR package to the NCDC Archive and make available for public access.

5.2 Production Phase

When a CDR enters the production phase, it is assumed to be stable and operationally efficient. New developments may trigger a desire to re-process the CDR. Since reprocessing is resource intensive, decisions about when and whether to reprocess a CDR will be based upon pre-determined criteria and with input from the scientific community. Where new satellite data are involved, the scientific community will determine if and when that new satellite data should be used to extend the previous record. The following steps comprise production:

5.2.1 Planning

Planning for production should begin during the transition of the CDR to FOC.

- a. Using statistics gathered during system integration and test, determine and make available appropriate production, data transfer, and storage environments.
- b. If the record is to be extended, re-evaluate the previously projected frequency with which the CDR will be extended and what resources are required for extension of the record.

5.2.2 Production

- a. Perform forward processing to extend the CDR or reprocess the CDR.
- b. Continue to gather performance and volume metrics with each new run.

5.2.3 Operational monitoring

- a. Utilize available automated tools as well as the scientific community to monitor output for issues and anomalies.
- b. When significant issues are encountered, engage personnel who understand and use the data for explanation or exploration of cause.

5.2.4 Preservation

- a. Prepare a new or updated Submission Agreement between the Operations Branch and the NCDC Archive for the CDR output.
- b. Transfer output data to the NCDC Archive and make available for public access.
- c. Maintain CDRs as per NCDC Archive standards.

5.2.5 Accessibility

- a. Provide data visualization tools as appropriate.
- b. Provide methods to subset data or to perform temporal or spatial searches of data.
- c. Provide CIRs as appropriate.
- d. Gather metrics on requests for CDR data and citations of CDR products in scientific and user literature.

5.2.6 Re-evaluation

- a. Periodically review operational performance and volume statistics.
- b. Ensure proper resourcing for anticipated production, data transfer, and storage needs.

If an algorithm is improved or if a new algorithm becomes available, then a decision against pre-determined criteria will be made whether to reprocess the CDR. If reprocessing is to occur with a new algorithm, then the CDR process returns to the transition phase, during which the new algorithm is tested and verified before the CDR processing with that algorithm moves into production. Potential triggers for reprocessing include but are not limited to the following:

- Algorithm improvements
- The discovery of processing anomalies
- The availability of better calibration data
- The availability of better ancillary data
- Calibration/normalization of new sensor data
- Better geophysical models

6. IOC to FOC Process Flow

Figure 1 illustrates the flow of the IOC to FOC process as described in Section 5 above.

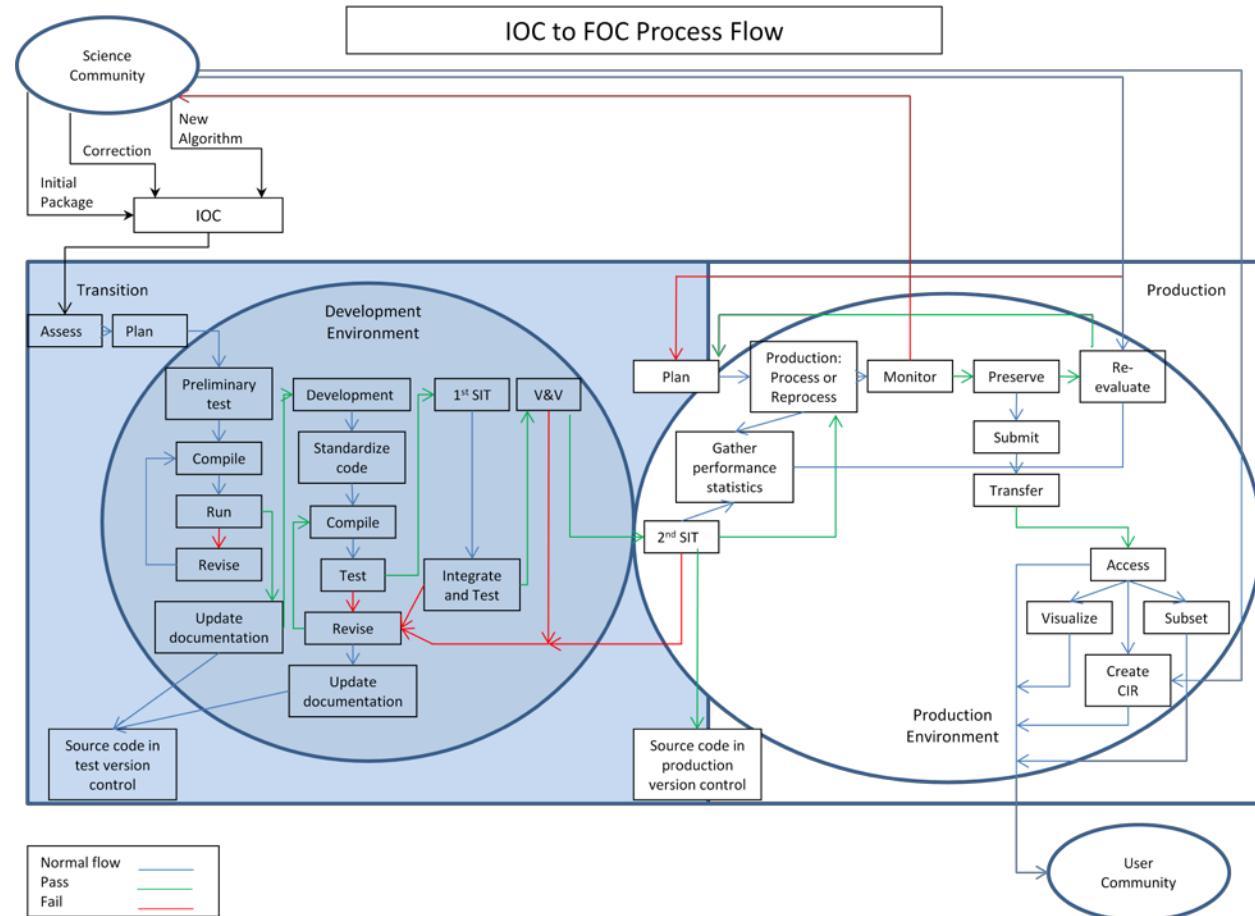


Figure 1: IOC to FOC Process Flow