Proposed CDR Definitions, etc.

Jeffrey L. Privette

Deputy Director, Center for Weather and Climate

National Centers for Environmental Information

Revealing the Past, Interpreting the Present, and Informing the Future

Rules Hierarchy

CDR development and generation rulebook needs:

1. CDR definitions (short, simple, general, unchanging)
2. CDR requirements (derived from definitions, unchanging, limited)
3. CDR production guidelines (specific, detailed, practical steps to meet requirements; evolve with technology, know-how, user needs and preferences)

Proposed Definitions

- An FCDR consists of a consistently-processed time series of uncertainty-quantified sensor observations calibrated to physical units, located in time and space, and of sufficient length and quality to be useful for climate science or applications.

- A CDR consists of a consistently-processed time series of uncertainty-quantified retrieved values of a geophysical variable or related indicator, located in time and space, and of sufficient length and quality to be useful for climate science or applications.

- An Interim Climate Data Record (ICDR) is consistently-processed times series of uncertainty-quantified estimates of CDR values produced at lower latency than, but otherwise minimizing differences with, the estimated CDR values.

Proposed Non-Functional Requirements*

A CDR compliant with GCOS science objectives is:

1. Extensible: Sustainable in time and with uninterrupted product generation.
2. Scientifically-defensible: The processing system adheres to commonly-accepted scientific principles and practices.
3. Transparent: All algorithms, processes and procedures are available for public inspection.
4. Reproducible: Processing system is well-documented and maintained under configuration control.
5. Sustainable: The inputs, processing and output architecture are designed to accommodate change.
6. Preserved: All data, documentation and source codes used for production, calibration and validation are preserved in a suitable archive.
7. Accessible: All data sets, documentation and source codes are publicly available.

*informed by various expert bodies, including U.S. NAS, GCOS, SCOPE-CM, WGClimate

Proposed Guidelines

Per previously introduced documentation

1. This CDR definition is intended to define the target objective (goal) for CDR activities. In the following points, “should” is used to indicate what is necessary to achieve this CDR definition. Importantly, initiatives that deliver societal and scientific benefit from reprocessing and improvement activities but meet this goal only partially should not for that reason be discouraged or inhibited.
2. A CDR is a record that is specifically useful for quantifying aspects of Earth’s climate system (but may also support non-climate applications).
3. The CDR may be provided at more than one level of processing: at the full resolution and with the geolocation of the underlying FCDR (level 2), gridded and/or merged (level 3), and/or gap-filled (level 4). The higher levels should be derived from and consistent with the lower-level CDR.
4. The CDR should cover a duration long enough to quantify climate variability and change. No single duration can be specified, since this depends on the variable, the uncertainty and stability of observation, and the climate science context. Usually, the label CDR would not be applied for satellite records shorter than 10 years. During periods of overlaps of sensors constituting a measurement series, all overlapping data should be included in the CDR at level 2, since overlaps should be exploited to maximise stability, including in derived higher-level CDRs.
5. The CDR should be stabilised, which means it should be harmonised or otherwise improved so as to maximise observational stability and/or meet climate requirements for observational stability.
6. The complete CDR should include unambiguous traceability to the underlying lower-level data (FCDRs) and the ancillary data from which geophysical values and their estimated uncertainty derive.
7. The data in the CDR should be located in time and space using a clearly defined co-ordinate system.
8. Quantitative uncertainty information should be provided with the CDR observations. Examples of uncertainty information include: standard uncertainty, fractional uncertainty, and error covariance matrices. Uncertainty may be provided as data arrays or (where valid, in order to minimise data volumes) as parametric expressions to calculate uncertainty from other data in the CDR. Uncertainty information should be provided per file or per datum as necessary to represent any significant variability of uncertainty, while taking account of data volume. Uncertainty estimates should cover all important sources of error, and be validated, documented and traceable.
9. Flags relating to assessment of data quality should be made available in the CDR together with the observations and their uncertainty.

Proposed CDR Guidelines (from WGC-10)

1. This CDR definition is intended to define the target objective (goal) for CDR activities. In the following points, “should” is used to indicate what is necessary to achieve this CDR definition. Importantly, initiatives that deliver societal and scientific benefit from reprocessing and improvement activities but meet this goal only partially should not for that reason be discouraged or inhibited.
2. A CDR is a record that is specifically useful for quantifying aspects of Earth’s climate system (but may also support non-climate applications).
3. The CDR may be provided at more than one level of processing: at the full resolution and with the geolocation of the underlying FCDR (level 2), gridded and/or merged (level 3), and/or gap-filled (level 4). The higher levels should be derived from and consistent with the lower-level CDR.
4. The CDR should cover a duration long enough to quantify climate variability and change. No single duration can be specified, since this depends on the variable, the uncertainty and stability of observation, and the climate science context. Usually, the label CDR would not be applied for satellite records shorter than 10 years. During periods of overlaps of sensors constituting a measurement series, all overlapping data should be included in the CDR at level 2, since overlaps should be exploited to maximise stability, including in derived higher-level CDRs.
5. The CDR should be stabilised, which means it should be harmonised or otherwise improved so as to maximise observational stability and/or meet climate requirements for observational stability.
6. The complete CDR should include unambiguous traceability to the underlying lower-level data (FCDRs) and the ancillary data from which geophysical values and their estimated uncertainty derive.
7. The data in the CDR should be located in time and space using a clearly defined co-ordinate system.
8. Quantitative uncertainty information should be provided with the CDR observations. Examples of uncertainty information include: standard uncertainty, fractional uncertainty, and error covariance matrices. Uncertainty may be provided as data arrays or (where valid, in order to minimise data volumes) as parametric expressions to calculate uncertainty from other data in the CDR. Uncertainty information should be provided per file or per datum as necessary to represent any significant variability of uncertainty, while taking account of data volume. Uncertainty estimates should cover all important sources of error, and be validated, documented and traceable.
9. Flags relating to assessment of data quality should be made available in the CDR together with the observations and their uncertainty.

Proposed FDR Guidelines (from WGC-10)

1. This FDR definition is intended to define the target objective (goal) for FDR activities. In the following points, “should” is used to indicate what is necessary to achieve this FDR definition. Importantly, initiatives that deliver societal and scientific benefit from reprocessing and improvement activities but meet this goal only partially should not for that reason be discouraged or inhibited. [Hard to understand]

2. The FDR should be consistently reprocessed or otherwise improved so as to maximise observational temporal consistency, in support of more consistent derived data records.

3. The complete FDR should include lower-level observations with the means to derive calibrated quantities from these (such as Earth view counts together with the appropriate parameters for transformation to physical units). This enables efficient re-estimation of calibrated quantities should further research improve understanding of the sensors’ calibration. Users may additionally require a consistent version of the FDR comprising calibrated measured values in their physical units, particularly if the conversion to calibrated values has some complexity. The FDR further includes all further ancillary data, telemetry, etc, used in the process of calibration (or in deriving calibration parameters). “Together with” does not necessarily mean “in the same file”, but the FDR should provide unambiguous traceability between the lower-level data, the calibrated data and the corresponding ancillary data.

4. The data in the FDR should be located in time and space using a clearly defined co-ordinate system, exploiting the best available estimates of orbital elements (where relevant).

5. Quantitative uncertainty information should be provided with the FDR observations. The form of uncertainty information should reflect the nature of the sensor and its error characteristics, as well as the requirements for quantifying uncertainty in derived higher-level data records. Examples of uncertainty information include: standard uncertainty, fractional uncertainty, and error covariance matrices. Uncertainty may be provided as data arrays or (where valid, in order to minimise data volumes) as parametric expressions to calculate uncertainty from other data in the FDR. Uncertainty information should be provided per orbit/file, per scan or per datum as necessary to represent any significant variability of uncertainty, while taking account of data volume. Uncertainty estimates should cover all important sources of error, and be validated, documented and traceable. Data underlying or ancillary to the uncertainty estimates should also be provided available together with the uncertainty information in the FDR.

6. Flags relating to assessment of data quality and instrument status should be made available in the FDR together with the observations and their uncertainty. To increase usability, summary flags may be provided that give simple guidance to users about the quality status of pixels based on expert understanding of which of the flags and instrument conditions indicate questionable or invalid data.

7. The FDR format should be consistent across the various sensors of the series. Versioning and informative metadata should be implemented, including links from calibrated observations to ancillary data (if not in the same files).

ided at more than one level of processing: at the full resolution and with the geolocation of the underlying FCDR (level 2), gridded and/or merged (level 3), and/or gap-filled (level 4). The higher levels should be derived from and consistent with the lower-level CDR.

Proposed FCDR Guidelines (from WGC-10)

1. This FCDR definition is intended to define a target objective (goal) for FCDR activities. In the following points, “should” is used to indicate what is necessary to achieve this FCDR definition. Importantly, initiatives that deliver societal and scientific benefit from reprocessing and improvement activities but meet this goal only partially should not for that reason be discouraged or inhibited.

2. An FCDR is a record that is specifically useful for quantifying aspects of Earth’s climate system (but may also support non-climate applications).

3. The FCDR should cover a duration long enough to quantify climate variability and change. No single duration can be specified, since this depends on the variable, the uncertainty and stability of observation, and the climate science context. Usually, the label FCDR would not be applied for satellite records shorter than 10 years. During periods of overlaps of sensors constituting a measurement series, all overlapping data should be included in the FCDR, since overlaps should be exploited for harmonisation and when deriving CDRs to maximise stability.

4. The FCDR should be stabilised, which means it should be harmonised or otherwise improved so as to maximise observational stability. Harmonisation is the recalibration of the sensor observations, using a stated reference and/or overlaps with other sensors in the series. The purpose of the recalibration is to bring consistency between sensors given the known (best estimate) differences in instrument characteristics (e.g., spectral response functions) between sensors. A harmonised (or otherwise stabilised) FCDR should enable more stable CDRs to be derived. The word stabilised in the definition should be considered to encompass this technical definition. Other approaches to stabilisation may be valid, such as homogenisation.

8.9. The complete FCDR should include lower-level observations with the means to derive calibrated quantities from these (such as Earth view counts together with the appropriate parameters for transformation to physical units). This enables efficient re-estimation of calibrated quantities should further research improve understanding of the sensors’ calibration. Users may additionally require a version of the FCDR comprising calibrated measured values in their physical units, particularly if the conversion to calibrated values has some complexity. In both cases, the observations are provided together with all further ancillary data, telemetry, etc, used in the process of calibration (or in deriving calibration parameters). “Together with” does not necessarily mean “in the same file”, but the FCDR should provide unambiguous traceability between the lower-level data, the calibrated data and the corresponding ancillary data.

5. The data in the FCDR should be located in time and space using a clearly defined co-ordinate system, exploiting the best available estimates of orbital elements (where relevant).

6. Quantitative uncertainty information should be provided with the FCDR observations. The form of uncertainty information should reflect the nature of the sensor and its error characteristics, as well as the requirements for quantifying uncertainty in derived CDRs. Examples of uncertainty information include: standard uncertainty, fractional uncertainty, and error covariance matrices. Uncertainty may be provided as data arrays or (where valid, in order to minimise data volumes) as parametric expressions to calculate uncertainty from other data in the FCDR. Uncertainty information should be provided per orbit/file, per scan or per datum as necessary to represent any significant variability of uncertainty, while taking account of data volume. Uncertainty estimates should cover all important sources of error, and be validated, documented and traceable. Data underlying or ancillary to the uncertainty estimates should also be provided together with the uncertainty information in the FCDR.

7. Flags relating to assessment of data quality and instrument status should be made available in the FCDR together with the observations and their uncertainty. To increase usability, summary flags may be provided that give simple guidance to users about the quality status of pixels based on expert understanding of which of the flags and instrument conditions indicate questionable or invalid data.