NOAA Technical Information Series NESDIS DSMR-00163 Version 1.0



doi: 10.25923/3702-5641

Data Stewardship Maturity Report for Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily)

Table 1 Legend						
Level 1	Level 2	Level 3	Level 4	Level 5		
Ad Hoc	Minimal	Intermediate	Advanced	Optimal		
Little or no management	Limited Management	Defined Management, partially implemented	Well-defined Management, fully implemented	Full Management, audited, measured, controlled		

Table 1. Scores for the Nine DSMM Key Components at a Glance					
Preservability - 4.5 Accessibility - 2 Usability - 4					
Production Sustainability - 4 Data Quality Assurance - 4		Data Quality Control/Monitoring - 2.5			
Data Quality Assessment - 3	Transparency/Traceability - 3.5	Data Integrity - 3.5			

NOAA National Centers for Environmental Information January 2020



U.S. DEPARTMENT OF COMMERCE

National Oceanic and Atmospheric Administration National Environmental Satellite, Data, and Information Service Cover Image: Data Stewardship Rating Diagram for Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily)

Shades of green are used to represent level 1 through level 5 ratings; denoting Ad Hoc, Minimal, Intermediate, Advanced, and Optimal stages for each of the nine key components, respectively. The dark green level indicates all the practices are completely satisfied. The lighter green levels indicate only some of the practices are satisfied. The lightest green level indicates none of the practices are satisfied.

The stewardship maturity of NCEI data product, Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily), is assessed based on a reference stewardship maturity framework. The current maturity ratings of Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily) are at Level 1 or higher for all nine key components with zero Level 1, two Level 2, three Level 3, four Level 4, and zero Level 5 key components.

NOAA Technical Information Series NESDIS DSMR-00163 Version 1.0

The National Environmental Satellite, Data, and Information Service (NESDIS) manages the Nation's civil Earth-observing satellite systems, as well as global national data bases for meteorology, oceanography, geophysics, and solar-terrestrial sciences. From these sources, it develops and disseminates environmental data and information products critical to the protection of life and property, national defense, and the national economy, energy development and distribution, global food supplies, and the development of natural resources.

Publication in the NOAA Technical Memorandum series does not preclude later publication in scientific journals in expanded or modified form. The NESDIS series of NOAA Technical Reports is a continuation of the former NESS and EDIS series of NOAA Technical Reports and the NESC and EDS series of Environmental Science Services Administration (ESSA) Technical Reports.

Copies of earlier reports may be available by contacting NESDIS Chief of Staff, NOAA/ NESDIS, 1335 East-West Highway, SSMC1, Silver Spring, MD 20910, (301) 713-3578.

ASSESSMENT REVISION HISTORY

Revision	Description	Date
V01r00	Initial Release	12/16/2021

NOAA Technical Information Series NESDIS DSMR-00163 Version 1.0

doi: 10.25923/3702-5641

Data Stewardship Maturity Report for Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily)

Lori Hager, Paul Lemieux III NOAA's National Centers of Environmental Information (NCEI) 151 Patton Avenue, Asheville, NC 28801, (828) 271-4800

Recommended Citation

Lori Hager, Paul Lemieux III. (2021), Data Stewardship Maturity Report for Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily)

NOAA Technical Information Series NESDIS DSMR-00163 Version 1.0, 21pp., doi: 10.25923/3702-5641

Table of Contents

List of Tables	7
Preface	8
1. Introduction	9
2. Results	10
3. Acknowledgment	15
4. References	16
Appendix I The Scientific Data Stewardship Maturity Matrix (DSMM)	18

List of Tables

Table 1. Scores for the Nine DSMM Key Components at a Glance	1
Table 2. Dataset and Data Stewardship Maturity Assessment Metadata	11
Table 3. Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM	
Key Components for the Dataset.	12

Preface

In response to the President's Open Government Initiative and related policies, NOAA has committed to providing improved public access to all of its environmental information, to enable research and commercial innovation through ease of data discovery and use [*Casey*, 2016].

OneStop supports NOAA's efforts by leveraging existing access technologies and infusing specific innovations to provide improved discover, access, and visualization services for NOAA's data. Also, OneStop is viewed by a NESDIS as a pathfinder effort with an initial focus on selected high-priority datasets from NESDIS and other program data meeting OneStop standards, but eventually scalable across NOAA's data. Lastly, OneStop is implementing the USGEO Common Framework for Earth Observation Data and leveraging/supporting the NOAA Big Data Project (BDP) and Big Earth Data Initiative (BEDI) [*Casey*, 2016].

As with any process of improvement planning, agencies need to find out where they are in terms of their compliance to the federal regulations and what they need to do if any areas of non-compliance are identified. To this end, a unified framework would be beneficial for assessing the current stage of stewardship practices applied to individual datasets and for providing a road map that will guide future investments towards enhanced stewardship of environmental datasets. The value and quality of a dataset depends in part on the stewardship practices applied after its development and production. Therefore, a unified framework providing a holistic view of the quality of stewardship practices applied to individual datasets is beneficial to data stewards and users [*Casey*, 2016].

The Data Stewardship Maturity Matrix (DSMM), jointly developed by domain (data management, technology, and science) subject matter experts from NOAA's National Centers for Environmental Information (NCEI) and Cooperative Institute for Climate and Satellites – North Carolina (CICS-NC), provides such a consistent framework [*Peng et al.*, 2016]. The DSMM, leveraging institutional knowledge and community practices and standards, defines a graduated maturity scale for each of nine key components of scientific data stewardship to enable a consistent assessment of the measureable stewardship practices applied to a given data set or product.

The NOAA Data Stewardship Maturity Technical Series captures stewardship maturity assessment results for individual datasets, provides consistent representation and citable documents of those assessments, ensures transparency, and allows better data quality information integration and content-based search and discovery of NOAA data.

NOAA Technical Information Series NESDIS DSMR-00163 Version 1.0

Data Stewardship Maturity Report for Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily)

1. Introduction

1.1 Purpose

The purpose of this document is to describe the results of stewardship maturity assessment for NOAA Climate Data Record for Mean Layer Temperature (Upper Troposphere & Lower Stratosphere from UCAR, Version 2, utilizing the Scientific Data Stewardship Maturity Matrix or DSMM [Peng, et al, 2016]. DSMM defines levels of stewardship maturity stages for Preservability, Accessibility, Usability, Production Sustainability, Data Quality Assurance, Data Quality Control/Monitoring, Data Quality Assessment, Transparency/Traceability, and Data Integrity key components. Each of these components is ranked from 'Ad hoc' to 'Optimal' (see Appendix I). This report is based on evaluation performed by NOAA OneStop metadata specialists working with Subject Matter Experts and utilizing the DSMM template [Peng, 2016].

1.2 Scope

Assessing stewardship maturity - the current state of how datasets are documented, preserved, stewarded, and made accessible publicly, is a critical step towards meeting U.S. federal regulations, organizational requirements, and user needs [Peng et al., 2016]. The goal of this document is to provide consistent and transparent stewardship maturity information to data users and decision-makers.

1.3 Dataset Abstract

The GPCP Daily analysis is a companion to the GPCP Monthly analysis, and provides globally complete precipitation estimates at a spatial resolution of one degree latitude-longitude and daily time scale from October 1996 to the present. Although derived using both some of the same, but also different, data sets and methods than used in the GPCP Monthly analysis, the GPCP Daily "adds up" to the GPCP Monthly. The GPCP Daily V1.3 analysis is currently computed by the University of Maryland and submitted to NCEI. The routine update of the product takes place two months after the end of the month, once all input data sets become available. The data set is part of World Climate Research Program (WCRP) and GEWEX activities, being part of the array of data sets describing the water and energy cycles of the planet under the auspices of the GEWEX Data and Assessment Panel (GDAP). Details of input data sets and methods can be found in: Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B McGavock, J. Susskind, 2001: Global Precipitation at One-Degree Daily Resolution from Multi-Satellite Observations. J. Hydrometeor., 2(1), 36-50.

1.4 Document Maintenance

This document is generated and maintained by NOAA's National Centers for Environmental Information. More on policy is available at https://www.ncei.noaa.gov/.

2. Results

The data stewardship maturity assessment information is summarized in Table 1. Each component is displayed along with its corresponding score in a color-coded table.

Table 2. Dataset and Data Stewardship Maturity Assessment Metadata				
Dataset Title	Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily)			
Dataset Information URL	https://doi.org/10.7289/V5RX998Z			
Data Provider POC (Name; Email; Affiliation)	Customer Engagement Branch, ncei.orders@noaa.gov, DOC/NOAA/NESDIS/NCEI > National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce			
Dataset POC (Name; Email; Affiliation)	NOAA Climate Data Record Program Office, gpcp_contacts@ noaa.gov, DOC/NOAA/NESDIS/NCEI > National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce			
SMM Version (Document ID and Version Number)	NCDC-CICS-SMM_0001_Rev.1 12/09/2014			
SMM POC (Name; E-mail; Affiliation)	Ge Peng, ge.peng@uah.edu, University of Alabama- Huntsville			
SMM Template Version (Document ID and Version Numbers)	NCDC-CICS-SMM_0001_Rev.1 v4.0 06/23/2015			
SMM Template POC	Ge Peng, ge.peng@uah.edu, University of Alabama- Huntsville			
SMM Assessment Version (v <nn>r<mm>, e.g., v01r00)</mm></nn>	v02r02			
SMM Assessment Date (MM/DD/YYYY)	03/14/2018			
SMM Assessment POC (Name; E-mail; Affiliation)	Paul Lemieux III, paul.lemieux@noaa.gov, Earth Resources Technology, Inc.			
Stewardship Maturity Ratings (each key component) (kc1/kc2/kc3/kc4/kc5/kc6/kc7/kc8/kc9)	4.5 / 2 / 4 / 4 / 2.5 / 3 / 3.5 / 3.5			
SMM Original Assessment Date (MM/DD/YYYY)	09/20/2017			
SMM Original Assessment POC (Name; E-mail; Affiliation)	Paul Lemieux III, paul.lemieux@noaa.gov, Earth Resources Technology, Inc.			
SMM Last Modified Date (MM/DD/YYYY)	10/15/2021			
SMM Last Modification POC (Name; E-mail; Affiliation)	Lori Hager, lori.hager@noaa.gov, CASE Consultants International			
SMM Modified Date (MM/DD/YYYY)	03/14/2018			
SMM Modification POC (Name; E-mail; Affiliation)	Paul Lemieux III, paul.lemieux@noaa.gov, Earth Resources Technology, Inc.			

Preservability	 Level 4.5 Archived at NOAA NCEI-NC. Following NOAA Climate Data Record (CDR) Research-2-Operation (R2O) transition process with the Initial Operation Capability (IOC). Following OAIS RM. Conforms to ISO 19115-2 metadata standard. Conforming to CDR Program (CDRP) guidelines on coding and NCEI Archive Branch (AB) guidance on file and variable naming conventions per Submission Agreement (SA). Plans to transition ISO metadata to newer 19115-1 standard. Comments: No known external audits of the archive performed at this time. Level 2 Collection level metadata is searchable online: https://data.noaa.gov/onestop/#/ Direct download via HTTPS: https://www.ncei.noaa.gov/data/global-precipitation-climatology-project-gpcp-daily/access/
	 Collection level metadata is searchable online: https://data.noaa.gov/onestop/#/ Direct download via HTTPS: https://www.ncei.noaa.gov/data/global-precipitation-
Accessibility	 Reports available internally for the FTP/HTTP & HAS servers. New technology for OneStop search and discovery planned (i.e. ElasticSearch, Hyrax Servers, etc.) This dataset is part of the CDR data group that will be OneStop ready. Comments: Dissemination reports are available internally, but not publicly. No granule/file level search capabilities.
Usability	 Level 4 NetCDF-4 data format (CF and ACDD compliant) is a community standard-based form with compliant metadata. Data Flow Diagram [Adler and NOAA CDR Program, 2017] is available online here: https://www.ncdc.noaa.gov/cdr/atmospheric/precipitation-gpcp-daily C-ATBD [Adler et al., 2017] is available online here: https://www.ncdc.noaa.gov/cdr/atmospheric/precipitation-gpcp-daily Comments: No known external rankings
Production Sustainability	Level 4 • Under NOAA CDR Operation & Maintenance (O&M) • Updated annually • Funding is allocated yearly • Product improvement process in place • CDR program under management by NCEI Comments:

DSMM Key Component	Stewardship Maturity Rating, Justification, and Comments
Data Quality Assurance	 Level 4 Agile development procedure in place with defined/fixed set of analysis metrics Master reference data are included in the source code package which is available online here: https://www.ncdc.noaa.gov/cdr/atmospheric/precipitation-gpcp-daily Quality flags available at the file level which qualifies as limited data quality assurance metadata
	Comments: No known external reviews
Data Quality Control/ Monitoring	 Level 2.5 DQC is done after each data processing Sampling and analysis is performed monthly with comparisons to climatologies but are not automatic. Procedures documented in the C-ATBD [Adler et al., 2017] available here: https://www.ncdc.noaa.gov/cdr/atmospheric/precipitation-gpcp-daily
	Comments: No data quality information in the metadata record
Data Quality Assessment	 Level 3 C-ATBD [Adler et al, 2017] is available online here: https://www.ncdc.noaa. gov/cdr/atmospheric/precipitation-gpcp-daily Research product assessed in literature [Huffman et al., 2009] available online here: https://doi.org/10.1029/2009GL040000 Several operational assessments in literature available Assessment carried out in the NCEI CDR R2O process CDR Initial Operational Capability (IOC) stage Product Maturity Matrix assessment [Adler and NOAA CDR Program, 2016] is available and online here: https://www.ncdc.noaa.gov/cdr/atmospheric/precipitation-gpcp-daily
	Comments: No data quality assessment information in the metadata record No known external ranking
Transparency / Traceability	 Level 3.5 CDR Program literature [Bates, Privette, Kearns, Glance, & Zhao, 2015] available online here: https://doi.org/10.1175/BAMS-D-15-00015.1 C-ATBD [Adler et al., 2017] available online here: https://www.ncdc.noaa. gov/cdr/atmospheric/precipitation-gpcp-daily DOI Assigned: 10.7289/V5RX998Z NCEI OID Assigned: DSI 3631_02 Dataset Configuration Management is EIA-649-B standard compliant and diagrammed this presentation document [Hutchins, 2015] available online here: https://www1.ncdc.noaa. gov/pub/data/sds/cdr/conferences/2015% 20PI% 20Annual% 20Meeting% 20-% 20Presentations/Day_1/(A-2)% 20Operations% 20and% 20Maintenance% 20(O_M)% 2 of% 20NOAA% 20IOC% 20CDRs% 20-% 20(Hutchins).pdf
	Comments: No OAD available

Table 3. Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the Dataset.				
DSMM Key Component	Stewardship Maturity Rating, Justification, and Comments			
Data Integrity	 Level 3.5 Checksums generated at ingest which verifies ingest integrity. Using standard-based technology for generating checksum at ingest. Checksum verified when customer orders data. Comments: No comments			

3. Acknowledgment

This work is supported by the NOAA OneStop Project.

We thank the dataset POCs for their valuable input, as well as the collaborative efforts of the OneStop teams, especially the Metadata team. We would also like to show appreciation to Ge Peng for her contributions.

The draft of this data stewardship maturity report is systematically generated by a tool created by Kieran Hodnett and populated with the stewardship maturity assessment done by the author(s) of this report. The tool was developed based on a Word template created collaboratively by Robert Partee II, Raisa Ionin, Paul Lemieux III, Ge Peng, Don Collins, and Sonny Zinn with helpful input from the NOAA Central Library and the NCEI Communication Team.

4. References

Casey, K. (2016), The NOAA OneStop data discover and access framework project, Version:June 3, 2016. https://cdn.ioos.noaa.gov/media/2017/12/OneStop-IOOS-DMAC-03-June-2016.pdf

Peng, G. (2015) The scientific data stewardship maturity assessment model template, Version: NCDC-CICS-SMM-0001-Rev.1 v4.0 6/23/2015. doi:10.6084/m9.figshare.1211954.

Peng, G., J.L. Privette, E.J. Kearns, N.A. Ritchey, and S. Ansari (2015), A unified framework for measuring stewardship practices applied to digital environmental datasets, *Data Science Journal*, 13, 231-253, doi: 10.2481/dsj.14-049.

Peng, G., J. Lawrimore, V. Toner, C. Lief, R. Baldwin, N. Ritchey, D. Brinegar, and S. A. Delgreco (2016) assessing stewardship naturity of the global historical climatology network-monthly (GHCN-M) dataset: use case study and lessons learned, D-Lib Magazine, 22, doi:10.1045/november2016-peng.

Adler, R., and NOAA CDR Program, (2017), GPCP Daily Analysis (V1.3) Data Flow Diagram, _Rep. CDRP-DIA-0914_, NOAA National Centers for Environmental Information, Asheville, NC., retrieved online: https://www.ncdc.noaa.gov/cdr/a tmospheric/precipitation-gpcp-daily/ (accessed 2017 Sept 20).

Huffman, G., Adler, R., Bolvin, D., and Gu, G., (2009), Improving the global precipitation record: GPCP Version 2.1, _Geophysical Research Letters_, 36(17), doi:10.1029/2009GL040000.

Adler, R., and NOAA CDR Program, (2016), Precipitation – GPCP Daily – NOAA Climate Data Record (CDR) Maturity Matrix, _Rep. CDRP-MM-0### Rev 1_, NOAA National Centers for Environmental Information, Asheville, NC., retrieved online: https://www.ncdc.noaa.gov/cdr/atmospheric/precipitation-gpcp-daily (accessed 2017 Sept 20).

Bates, J., Privette, J., Kearns, E., Glance, W., and Zhao, X. (2015), Sustained production of multidecadal climate records: lessons from the NOAA Climate Data Record Program, _Bulletin of the American Meteorological Society_, 97(10), 1573—1582, doi:10.1175/BAMS-D-15-00015.1

Adler, R., Sapiano, M., Wang, J. (2017), Climate Algorithm Theoretical Basis Document (C-ATBD) Precipitation – Global Precipitation Climatology Project (GPCP) Daily Analysis, _Rep. CDRP-ATBD-0913_, NOAA National Centers for Environmental Information, Asheville, NC. retrieved online: https://www.ncdc .noaa.gov/cdr/atmospheric/precipitation-gpcp-daily (accessed 20 Sept 2017 Hutchins, C. (2015), Operations and Maintenance (O&M) of NOAA IOC CDRs, https: //www1.ncdc.noaa.gov/pub/data/sds/cdr/conferences/2015%20PI%20Annual%20Mee ting%20-%20Presentations/Day_1/(A-2)%20Operations%20and%20Maintenance%20(O _M)%20of%20NOAA%20IOC%20CDRs%20-%20(Hutchins).pdf

Appendix I: The Scientific Data Stewardship Maturity Matrix (DSMM)

Table A1: This matrix (Version: NCDC-CICS-SMM-0001-Rev.1. 12/09/2014) describes the criterion used to evaluate data stewardship maturity for each of the nine DSMM key components [*Peng et al.*, 2015].

DSMM Component	Level 1 Ad hoc Little or no management	Level 2 <i>Minimal</i> Limited management	Level 3 <i>Intermediate</i> Defined management, partially implemented	Level 4 Advanced Well-defined management, fully implemented	Level 5 <i>Optimal</i> Full management, audited, measured, controlled
Preservability (The state of being preservable)	Any storage location Data only	Non- designated repository Redundancy Limited archiving metadata	Designated archive Redundancy Community- standard archiving metadata Conforming to limited archiving standards	Level 3 + Conforming to community archiving standards	Level 4 + Archiving process performance controlled, measured, and audited Future archiving standard changes planned
<i>Accessibility</i> (The state of being searchable and accessible publicly)	Not publically available person-to- person	Publically available direct file download (e.g., via anonymous FTP server) Collection or dataset level searchable online	Level 2 + Non-standard data service Limited data server performance Granule/file level searchable Limited search metrics	Level 3 + Community- standard data service Enhanced data server performance Conforming to community search metrics Dissemination report metrics defined and implemented internally	Level 4 + Dissemination reports available online Future technology and standard changes planned

Usability (The state of being easy to use)	Extensive product-specific knowledge required No documentation online	Non-standard data format Limited documentation (e.g., user's guide online)	Community standard-based interoperable format & metadata Documentation (e.g. source code, product algorithm document, processing or/and data flow diagram) online	Level 3 + Basic capability (e.g., subsetting, aggregating) & data characterization overall/global, e.g., climatology, error estimates) available online	Level 4 + Enhanced online capability (e.g., visualization, multiple data formats) Community metrics of data characterization (regional/cell) online External ranking
Production Sustainability (The state of data production being sustainable and extendable)	Ad Hoc or Not applicable To obligation or deliverable requirement	Short-term Individual PI's commitment (grant obligations)	Medium-term Institutional commitment (contractual deliverables with specs and schedule defined)	Long-term Institutional commitment Product improvement process in place	Level 4 + National or international commitment Changes for echnology planned
Data Quality Assurance (The state of data quality being assured)	Data quality assurance (DQA) procedure unknown or none	Ad Hoc and random QA procedure not defined and documented	DQA procedure defined and documented and partially implemented	DQA procedure well documented, fully implemented and available online with master reference data Limited data quality assurance metadata	Level 4 + DQA procedure monitored and reported Conforming to community quality metadata & standards External review

Data Quality Control/ Monitoring The state of data quality being controlled and monitored	None or Sampling unknown or spotty Analysis unknown or random in time	Sampling and analysis are regular in time and space Limited product-specific metrics defined & implemented	Level 2 + Sampling and analysis are frequent and systematic but not automatic Community metrics defined and partially implemented Procedure documented and available online	Level 3 + Anomaly detection procedure well-documente d and fully implemented using community metrics, automatic, tracked and reported Limited quality monitoring metadata	Level 4 + Cross-validation of temporal & spatial characteristics Physical consistency check Conforming to community quality metadata & standards
Data Quality Assessment (The state of data quality being assessed)	Algorithm/ method/model Theoretical basis assessed (methods and results online)	Level 1 + Research product assessed (methods and results online)	Level 2 + Operational product assessed (methods and results online)	Level 3 + Quality metadata assessed Limited quality assessment metadata	Level 4 + Assessment performed on a recurring basis Conforming to community quality metadata & standards External ranking
Transparency/ Traceability (The state of being transparent, trackable, and traceable)	Limited product information available Person-to- person	Product information available in literature	Algorithm Theoretical Basis Document (ATBD) & source code online Dataset configuration managed (CM) Unique Object Identifier (OID) assigned (dataset, documentation, source code) Data citation tracked (e.g., utilizing Digital Object Identifier	Level 3 + Operational Algorithm Description (OAD) online, OID assigned, and under CM	Level 4 + System information online Complete data provenance online

Data Integrity (The state of data integrity being verifiable)	Unknown or no data ingest integrity check	Data ingest integrity verifiable (e.g, checksum technology)	(DOI) system) Level 2 + Data archive integrity verifiable	Level 3 + Data access integrity verifiable Conforming to	Level 4 + Data authenticity verifiable (e.g., data signature technology)
			vennaoie	Conforming to community data integrity technology standard	-