**NOAA Technical Report**

**NESDIS GHCN-M-v3**

doi: 10.7289/XXXXXXX

**Data Stewardship Maturity Report for Global Historical Climatology Network-Monthly, Version 3**

NOAA National Centers for Environmental Information

December 2016

**U.S. DEPARTMENT OF COMMERCE**

**National Oceanic and Atmospheric Administration**

National Environmental Satellite, Data, and Information Service

**Cover Image: Generic mosaic image related to the dataset covered in this report (Content Editor could provide).**

**NOAA TECHNICAL MEMORANDUM SERIES**

**National Environmental Satellite, Data, and Information Service**

**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, glo0bal 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.**

**NOAA Data Stewardship Maturity Technical Series**

doi: 10.7289/XXXXXXXX

**Data Stewardship Maturity Report for**

**Global Historical Climatology Network-Monthly, Version 3**

**Valerie Toner, valerie.toner@noaa.gov, NCEI/DSD/AB**

NOAA’s National Centers of Environmental Information (NCEI), 151 Patton Avenue, Asheville, NC 28801-5001, USA

**Recommended Citation:**

Ionin, R., G. Peng, and K. Saha (2016), Data stewardship maturity report for GHRSST Level 4 AVHRR\_AMSR\_OI Global Blended Sea Surface Temperature Analysis (GDS Version 1), *NOAA/NESDIS Technical Report XXX*, 30, doi: 10.7289/XXXXXX.

**Contents**

List of Figures ……………………………………………………………………………………4

List of Tables ……………………………………………………………………………………..5

Preface ……………………………………………………………………………………………7

1. Introduction ………………………………………………………………………………9

2. Results …………………………………………………………………………………..10

3. Summary………………………………………………………………………………...23

4. Acknowledgments ……………………………………………………………………...23

5. References………………………….……….…….…………………………………….24

Appendix I………………………………………………………………………………………25

**List of Figures**

Figure 1. (Page X) Data stewardship maturity scoreboard for GHCN-M-v3, highlighted with 5-level progressive green scales for each of the nine key components (across), representing Ad Hoc, Minimal, Intermediate, Advance, and Optimal stages (vertical). If more than two cells are highlighted, it denotes that the dataset has completely satisfied the criterion for the lower level but not yet so at the current level.

Figure 2. (Page X) Data stewardship rating diagram for GHCN-M-v3. One to five stars are used to represent Level 1 to 5 ratings, denoting Ad Hoc, Minimal, Intermediate, Advance, and Optimal stages for each of the nine key components, respectively. The dark filled stars indicate that all the practices are completely satisfied. The light filled ones indicated that not all the practices are satisfied. And the non-filled ones indicated that the practices are not satisfied*.*

**List of Tables**

Table 1. (Page X) Dataset and Data Stewardship Maturity Assessment Metadata.

Table 2. (Page X) Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the GHCN-M-v3 Dataset

**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 goal of defining the stewardship maturity matrix for data stewards is to provide a holistic, consistent, quantifiable, and scalable measure of stewardship maturity for data users and stakeholders including data providers and decision-support system users. It is our hope that this undertaking will help alleviate the burden that data stewards face in defining their own assessment models and reduce incompatibility of stewardship maturity assessment results from individually defined models. Effort was taken to generalize the maturity levels of this stewardship assessment model to be applicable to diverse digital environmental data products in various scientific and user communities. The underlying best practices and standards in the stewardship maturity matrix are intended to be community-accepted to allow the flexibility of its implementation [*Peng et al.*, 2015].

**NOAA Data Stewardship Maturity Technical Series – 5/18/2016 14:45:58**

**Data Stewardship Maturity Report for**

**Global Historical Climatology Network-Monthly, Version 3**

**1. Introduction**

**1.1 Purpose**

The purpose of this document is to describe the results of stewardship maturity assessment for Global Historical Climatology Network-Monthly, Version 3, utilizing the Scientific Data Stewardship Maturity Matrix or *DSMM* [*Peng, et al*, 2016]. DSMM defines 5 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 are 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*, 2015].

**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 the consistent and transparent stewardship maturity information to data users and decision-makers.

**1.3 Dataset Outline**

A Group for High Resolution Sea Surface Temperature (GHRSST) global Level 4 sea surface temperature analysis produced daily on a 0.25 degree grid at the NOAA National Climatic Data Center. This product uses optimal interpolation (OI) using data from the Advanced Very High Resolution Radiometer (AVHRR) Pathfinder Version 5 time series (when available, otherwise operational NOAA AVHRR data are used), the Advanced Microwave Scanning Radiometer-EOS (AMSR-E), and in situ ship and buoy observations. A second similar product is available back to 1981 that includes only in situ and AVHRR Pathfinder data in its analysis. The OI analysis is a daily average SST that is bias adjusted using a spatially smoothed 7-day in situ SST average and is thus tuned to about 0.3 meter. Both day and night satellite fields are independently bias adjusted. More information is available at <https://www.ncdc.noaa.gov/oisst>.

**1.4 Document Maintenance**

This document is generated and maintained by NOAA’s National Centers for Environmental Information. [More on policy here](https://www.ngdc.noaa.gov/).

**2. Results**

The information about dataset and stewardship maturity assessment is summarized in Table 1. The data stewardship maturity ratings are displayed as the scoreboard (Figure 1)and rating diagram (Figure 2) with the detailed justifications in Table 2.

Table 1. Dataset and Data Stewardship Maturity Assessment Metadata.

|  |  |
| --- | --- |
| **Dataset Title** | Global Historical Climatology Network-Monthly, Version 3 |
| **Dataset Information URL** | http://www.ncdc.noaa.gov/ghcnm/v3.php; http://dx.doi.org/10.7289/V5X34VDR; |
| **Data Provider POC (Name; E-mail: Affiliation)** | Jay Lawrimore, Jay.Lawrimore@noaa.gov, NCEI/CWC/CSD/DSB |
| **Dataset POC (Name; E-mail; Affiliation)** | Jay Lawrimore, Jay.Lawrimore@noaa.gov, NCEI/CWC/CSD/DSB |
| **SMM Version (Document ID and Version Number)** | NCDC-CICS-SMM\_001\_Rev.1 |
| **SMM POC (Name; E-mail; Affiliation)** | Ge Peng, Ge Peng@noaa.gov, CICS-NC/NCEI |
| **SMM Template Version (Document ID and Version Numbers)** | NCDC\_CICS\_SMM\_0001\_Rev1\_template\_v4.0\_20150623 |
| **SMM Template POC** | Jay Lawrimore;,Jay.Lawrimore@noaa.gov, NCEI/CWC/CSD/DSB; Valerie Toner, valerie.toner@noaa.gov, NCEI/DSD/AB; Christina Lief, Christina.Lief@noaa.gov, NCEI/DSD/AB; Ge Peng, Ge.Peng@noaa.gov, NCEI/CWC/CSD/PRB & CICS-NC; Rich Baldwin, Rich.Baldwin@noaa.gov,NCEI/DSD/DAB |
| **SMM Assessment Version (v<nn>r<mm>, e.g., v01r00)** | v01r03 |
| **SMM Assessment Date (MM/DD/YYYY)** | 2015-09-04 |
| **SMM Assessment POC (Name; E-mail; Affiliation)** | Valerie Toner, valerie.toner@noaa.gov, NCEI/DSD/AB |
| **Stewardship Maturity Ratings (each key component) (kc1/kc2/kc3/kc4/kc5/kc6/kc7/kc8/kc9)** | 4/2/2.5/4.5/3.5/2.5/3/2.5/3.5 |
| **SMM Original Assessment Date (MM/DD/YYYY)** | 2015-06-08 |
| **SMM Original Assessment POC (Name; E-mail; Affiliation)** | Valerie Toner, valerie.toner@noaa.gov, NCEI/DSD/AB |
| **SMM Last Modified Date (MM/DD/YYYY)** | N/A |
| **SMM Last Modification POC (Name; E-mail; Affiliation)** | N/A |
| **SMM modified Date (MM/DD/YYYY)** | N/A |
| **SMM Modification POC (Name; E-mail; Affiliation)** | N/A |

Table 2. Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the GHCN-M-v3 Dataset.

Table 1. Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the GHCN-M-v3 Dataset.

| **DSMM Key Component** | **Stewardship Maturity Rating, Justification, and Comments** |
| --- | --- |
| ***Preservabilty*** *(The state of being preservable)* | **★ Level 4**  ***Justification:***  N/A  ***Comments:***  N/A |
| ***Accessibility***  *(The state of being searchable and accessible publically)* | **★ Level 2**  ***Justification:***  N/A  ***Comments:***  N/A |
| ***Usability***  *(The state of being easy to use)* | **★ Level 2.5**  ***Justification:***  N/A  ***Comments:***  N/A |
| ***Production Sustainability*** *(The state of data production being sustainable and extendable)* | **★ Level 4.5**  ***Justification:***  N/A  ***Comments:***  N/A |
| ***Data Quality Assurance*** *(The state of data quality being assured)* | **★ Level 3.5**  **Justification:**  N/A  ***Comments:***  N/A |
| ***Data Quality Control/Monitoring*** *(The state of data quality being controlled and monitored)* | **★ Level 2.5**  ***Justification:***  N/A  ***Comments:***  N/A |
| ***Data Quality Assessment*** *(The state of data quality being assessed)* | **★ Level 3**  ***Justification:***  N/A  ***Comments:***  N/A |
| ***Transparency*** *(The state of being transparent, trackable, and traceable)* | **★ Level 2.5**  ***Justification:***  N/A  ***Comments:***  N/A |
| ***Data Integrity*** *(The state of data integrity being verifiable)* | **★ Level 3.5**  ***Justification:***    N/A  ***Comments:***    N/A |



Figure 1. Data stewardship maturity scoreboard for GHCN-M-v3, highlighted with 5-level progressive green scales for each of the nine key components (across), representing Ad Hoc, Minimal, Intermediate, Advance, and Optimal stages (vertical). If more than two cells are highlighted, it denotes that the dataset has completely satisfied the criterion for the lower level but not yet so at the current level.



Figure 2. Data stewardship rating diagram for GHCN-M-v3. One to five stars are used to represent Level 1 to 5 ratings, denoting Ad Hoc, Minimal, Intermediate, Advance, and Optimal stages for each of the nine key components, respectively. The dark filled stars indicate that all the practices are completely satisfied. The light filled ones indicated that not all the practices are satisfied. And the non-filled ones indicated that the practices are not satisfied**.**

**3. Summary**

The stewardship maturity of NCEI data product, GHCN-M-v3, is assessed based on a reference stewardship maturity framework. The current maturity ratings of GHCN-M-v3 are at Level <minimum of all> or higher for all nine key components with <NN> Level 2, <NN> Level 3, and <NN> Level 4 key components.

**4. Acknowledgment**

This work is supported by NOAA OneStop Project. We thank beneficial input from dataset POC(s) and collaborative effort by OneStop Teams, especially the Metadata Team. Guidance from Ge Peng on DSMM was beneficial.

**5. References**

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: <http://dx.doi.org/10.2481/dsj.14-049>.

Casey, K. (2016), The NOAA *OneStop* data discover and access framework project, <https://ioos.noaa.gov/wp-content/uploads/2016/06/OneStop-IOOS-DMAC-03-June-2016.pdf>.

[The reference section will be created automatically including the ones listed here.]

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. Lawrimore, V. Toner, C. Lief , R. Baldwin, N. Ritchey, D. Brinegar, and S. A. Delgreco (2016) Assessing Stewardship Maturity 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.

**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 Maturity Component** | **Level 1 *Ad hoc***  Little or no management | **Level 2**  ***Minimal***  Limited management | **Level 3**  ***Intermediate***  Defined management, partially implemented | **Level 4**  ***Advanced***  Well-defined management, fully implemented | **Level 5**  ***Optimal***  Full management, audited, measured, controlled |
| --- | --- | --- | --- | --- | --- |
| ***Preservability*** *(The state of being perservable)* | 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 |
| ***Accessibility***  *(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-documented 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 (DOI) system) | 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) | Level 2 +  Data archive integrity verifiable | Level 3 +  Data access integrity verifiable  Conforming to community data integrity technology standard | Level 4 +  Data authenticity verifiable (e.g., data signature technology)  Performance of data integrity check monitored and reported |