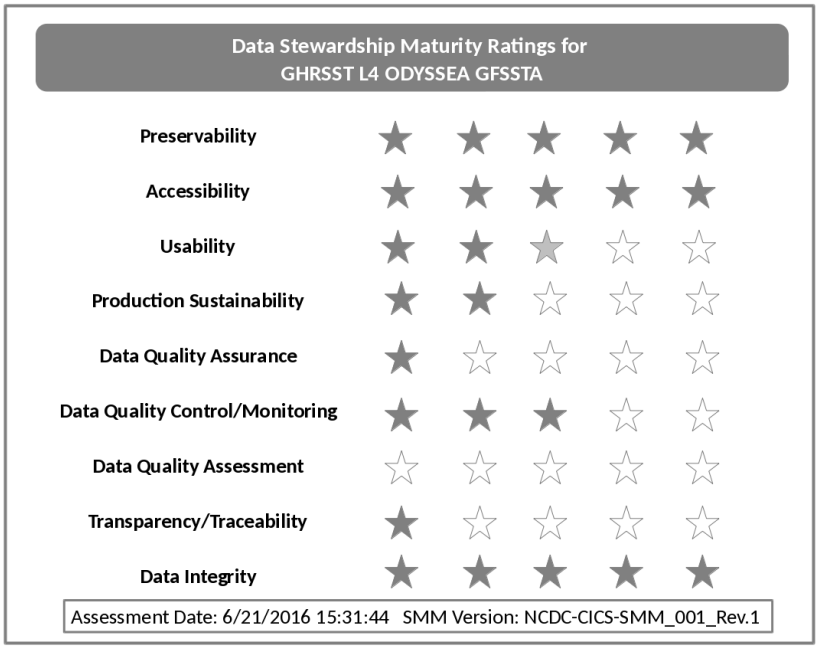
NOAA Technical Information Series NESDIS XXX
Version 1.0

doi: 10.7289/XXXXXXX

Data Stewardship Maturity Report for GHRSST Level 4 ODYSSEA Global Foundation Sea Surface Temperature Analysis (GDS version 1)



Dark solid filled stars – completely satisfied
Light solid filled stars – partially satisfied
Non-filled stars – not satisfied

NOAA National Centers for Environmental Information
6/21/2016

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Environmental Satellite, Data, and Information Service

Cover Image: Data stewardship rating diagram for GHRSST L4 ODYSSEA GFSSTA. 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**.**

The stewardship maturity of NCEI data product, GHRSST L4 ODYSSEA GFSSTA, is assessed based on a reference stewardship maturity framework.

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, 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.

NOAA Technical Information Series NESDIS XXX
Version 1.0
doi: 10.7289/XXXXXXXX

Data Stewardship Maturity Report for

GHRSST Level 4 ODYSSEA Global Foundation Sea Surface Temperature Analysis (GDS version 1)

N/A
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 ODYSSEA Global Foundation Sea Surface Temperature Analysis (GDS version 1), *NOAA/NESDIS Technical Report XXX*, 30 pp., doi: 10.7289/XXXXXX.

Contents

List of Figures ……6
List of Tables ……..7
Preface ……8
1. Introduction ……9
2. Results ……9
3. Summary…….....2
4. Acknowledgment ……...…16
5. References……..……..……..……...17

List of Figures

Figure 1. (Page 1) Data stewardship maturity scoreboard for GHRSST L4 ODYSSEA GFSSTA, 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.

List of Tables

Table 1. (Page 9) Dataset and Data Stewardship Maturity Assessment Metadata.

Table 2. (Page 19) Stewardship Maturity Levels and Detailed Justifications for Each of Nine DSMM Key Components for the GHRSST L4 ODYSSEA GFSSTA 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 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.*, 2015]. 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 Report NESDIS XXX

Data Stewardship Maturity Report for GHRSST Level 4 ODYSSEA Global Foundation Sea Surface Temperature Analysis (GDS version 1)

1. Introduction

1.1 Purpose

The purpose of this document is to describe the results of stewardship maturity assessment for GHRSST Level 4 ODYSSEA Global Foundation Sea Surface Temperature Analysis (GDS version 1), 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 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*, 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) Level 4 sea surface temperature analysis produced daily on an operational basis at Ifremer/CERSAT (France) using optimal interpolation (OI) on a global 0.1 degree grid. It provides a daily cloud-free field of foundation sea surface temperature at approximately 10km resolution (0.1 degree) over the full globe. It is generated by merging microwave and infrared satellite sea surface temperature observations including those from the Advanced Very High Resolution Radiometer (AVHRR), the Advanced Along Track Scanning Radiometer (AATSR), the Spinning Enhanced Visible and Infrared Imager (SEVIRI), the Advanced Microwave Scanning Radiometer-EOS (AMSRE), the Tropical Rainfall Measuring Mission Microwave Imager (TMI) and the Geostationary Operational Environmental Satellite (GOES) Imager. The satellite SST observations are intercalibrated using the AATSR sensor as a reference (previously re-calibrated using all available in situ data). The development of the global real-time sea surface temperature at Ifremer/CERSAT is supported by European Commission initially in the frame of MERSEA project.

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.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	GHRSST Level 4 ODYSSEA Global Foundation Sea Surface Temperature Analysis (GDS version 1)
Dataset Information URL	http://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:GHRSST-EUR-L4HRfnd-GLOB-ODYSSEA
Data Provider POC (Name; E-mail: Affiliation)	NCEI, NOAA, ncei.info@noaa.gov
Dataset POC (Name; E-mail; Affiliation)	Jean-Francois Piolle, jfpiolle@ifremer.fr, IFREMER/CERSAT
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	Raisa Ionin, raisa.ionin@noaa.gov, NOAA, NCEI
SMM Assessment Version (v<nn>r<mm>, e.g., v01r00)	V01r02
SMM Assessment Date (MM/DD/YYYY)	2016-06-21
SMM Assessment POC (Name; E-mail; Affiliation)	
Stewardship Maturity Ratings (each key component) (kc1/kc2/kc3/kc4/kc5/kc6/kc7/kc8/kc9)	5/5/2.5/2/1/3/0/1/5
SMM Original Assessment Date	2016-05-25

(MM/DD/YYYY)	
SMM Original Assessment POC (Name; E-mail; Affiliation)	
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 GHRSST L4 ODYSSEA GFSSTA Dataset.

DSMM Key Component	Stewardship Maturity Rating, Justification, and Comments
Preservability *(The state of being preservable)*	**★ Level 5** ***Justification:*** N/A ***Comments:*** N/A
Accessibility *(The state of being searchable and accessible publically)*	**★ Level 5** ***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 2** ***Justification:*** N/A ***Comments:*** N/A

Data Quality Assurance *(The state of data quality being assured)*	★ **Level 1** **Justification:** N/A ***Comments:*** N/A
Data Quality Control/Monitoring *(The state of data quality being controlled and monitored)*	★ **Level 3** ***Justification:*** N/A ***Comments:*** N/A

Data Quality Assessment *(The state of data quality being assessed)*	★ **Level 0** ***Justification:*** N/A ***Comments:*** N/A
Transparency *(The state of being transparent, trackable, and traceable)*	★ **Level 1** ***Justification:*** N/A ***Comments:*** N/A
Data Integrity *(The state of data integrity being verifiable)*	★ **Level 5** ***Justification:*** N/A ***Comments:*** N/A

SMM Document ID: GHRSST L4 ODYSSEA GFSSTA
Version: NCDC-CICS-SMM_001_Rev.1

GHRSST L4 ODYSSEA GFSSTA

Data Stewardship Maturity Scoreboard

Maturity Level as of 6/21/2016 15:31:44

Maturity Scale	Preservability	Accessibility	Usability	Production Sustainability	Data Quality Assurance	Data Quality Control/Monitoring	Data Quality Assessment	Transparency /Traceability	Data Integrity
Level 1 – Ad Hoc Not Managed	Any storage location Data only	Not publicly available Person-to-person	Extensive product-specific knowledge required No documentation online	Ad Hoc or Not applicable No obligation or deliverable requirement	Data quality assurance (DQA) procedure unknown or none	None or Sampling unknown or spotty Analysis unknown or random in time	Algorithm/method/model theoretical basis assessed (method and results online)	Limited product information available Person-to-person	Unknown or no data ingest integrity check
Level 2 - Minimal Managed Limited	Non-designated repository Redundancy Limited archiving metadata	Publicly available Direct file download (e.g., via anonymous FTP server) Collection/dataset level searchable	Non-standard data format Limited documentation (e.g., user's guide) online	Short-term Individual PI's commitment (grant obligations)	Ad Hoc and random DQA procedure not defined and documented	Sampling and analysis are regular in time and space Limited product-specific metrics defined & implemented	Level 1 + Research product assessed (method and results online)	Product information available in literature	Data ingest integrity verifiable (e.g., checksum technology)
Level 3 - Intermediate Managed Defined, Partially Implemented	Designated archive Redundancy Community-standard archiving metadata Conforming to limited archiving process standards	Level 2 + Non-standard data service Limited data server performance Granule/file level searchable Limited search metrics	Community Standard-based interoperable format & metadata Documentation (e.g., source code, product algorithm document, processing or/and data flow diagram) online	Medium-term Institutional commitment (contractual deliverables with specs and schedule defined)	DQA procedure defined and documented and partially 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 2 + Operational product assessed (method and results online)	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 2 + Data archive integrity verifiable
Level 4 - Advanced Managed Well-Defined, Fully Implemented	Level 3 + Conforming to community archiving standards	Level 3 + Community-standard data services Enhanced data server performance Conforming to community search metrics Dissemination report metrics defined and implemented internally	Level 3 + Basic capability (e.g., subsetting, aggregating) & data characterization (overall/global, e.g., climatology, error estimates) available online	Long-term Institutional commitment Product improvement process in place	DQA procedure well documented, fully implemented and available online with master reference data Limited data quality assurance metadata	Level 3 + Anomaly detection procedure well-documented and fully implemented using community metrics, automatic, tracked and reported Limited quality monitoring metadata	Level 3 + Quality metadata assessed (method and results online) Limited quality assessment metadata	Level 3 + Operational Algorithm Description (OAD) online, OID assigned, and under CM	Level 3 + Data access integrity verifiable Conforming to community data integrity technology standard
Level 5 - Optimal Level 4 + Measured , Controlled , Audit	Level 4 + Archiving process performance controlled, measured, and audited Future archiving standard changes planned	Level 4 + Dissemination reports available online Future technology and standard changes planned	Level 4 + Enhanced online capability (e.g., visualization, multiple data formats) Community metrics of data characterization (regional/cell) online External ranking	Level 4 + National or international commitment Changes for technology planned	Level 4 + DQA procedure monitored and reported Conforming to community quality metadata & standards External review	Level 4 + Cross-validation of temporal & spatial characteristics Physical consistency check Conforming to community quality metadata & standards Dynamic providers/users feedback in place	Level 4 + Assessment performed on a recurring basis Conforming to community quality metadata & standards External ranking	Level 4 + System information online Complete data provenance available online	Level 4 + Data authenticity verifiable (e.g., data signature technology) Performance of data integrity check monitored and reported

Dataset Information: **http://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:GHRSST-EUR-L4HRfnd-GLOB-ODYSSEA**
Dataset POC: **Jean-Francois Piolle, jfpiolle@ifremer.fr,**
IFREMER/CERSAT

SMM POC: **Ge Peng; Ge.Peng@noaa.gov**
SMM Assessment POC: **Raisa Ionin, raisa.ionin@noaa.gov, NOAA, NCEI**

Figure 1. Data stewardship maturity scoreboard for GHRSST L4 ODYSSEA GFSSTA, 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.

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.

The draft of this data stewardship maturity report is systematically generated by a tool created by Sonny Zinn, 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, Donald Collins, and Sonny Zinn with beneficial input from NOAA Central Library and NCEI Communication Team.

5. References

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 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 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
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	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

			diagram) 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/	Limited product	Product	Algorithm	Level 3 +	Level 4 +

Traceability *(The state of being transparent, trackable, and traceable)*	information available Person-to-person	information available in literature	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)	Operational Algorithm Description (OAD) online, OID assigned, and under CM	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