A Quick Startup Guide for Utilizing the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix

Ge Peng (Ge.Peng@noaa.gov)
CICS-NC/NCEI
Version: v02r00-External 20170714

Revision History

Revision	Author	Description	Date
v01r00	Ge Peng, Research Scholar, CICS-NC/NCEI	Initial release	06/17/2016
v01r011	Ge Peng	Added this "Revision History" table; Changes made to Section 4, including the addition of the URL for the DSMM graphics generator; Changes made to Section 5 to improve readability.	6/24/2016
v02r00-External	Ge Peng	Changes made to Sections 1, 4 and 5, including updated links to the DSMM guide and Use Case Submission Google Form for the non-NOAA users.	7/14/2017

1. Introduction

1.1 Purpose and Scope

The purpose of this document is to provide high-level background information on the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix (DSMM, hereinafter) and a step-by-step quick start-up guide on how to get the DSMM template, collect needed information, carry out stewardship maturity assessment of the dataset, and display consistently the assessment results. The section on integration to NCEI and NOAA metadata resources, i.e., Section 5, is still evolving and will be frequently updated.

1.2 What is the DSMM?

The DSMM is a unified framework for measuring stewardship practices applied to individual digital environmental datasets. DSMM is jointly developed by an integrated team of subject matter experts of multi-domains from Cooperative Institute for Climate and Satellite – North Carolina (CICS-NC), North Carolina State University and NOAA's National Centers for Environmental Information (NCEI), leveraging institutional knowledge and community best practices and standards.

There are nine quasi-independent key components in DSMM: Preservability, Accessibility, Usability, Production Sustainability, Data Quality Assurance, Data Quality Control/Monitoring, Data Quality Assessment, Transparency/Traceability, and Data Integrity. A five-level progressive maturity scale is defined for each component associated with measurable practices

applied to individual datasets, representing Ad Hoc, Minimal, Intermediate, Advanced, and Optimal stages (Figure 1, also see Peng et al. (2015) for the rationale for each key component and its maturity levels, available online at: http://tinyurl.com/DSMMpaper).

1.3 Referencing this document

This document should be referenced as follows:

Peng, G., 2017: A Quick Startup Guide for utilizing the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix - External. Version: v02r00-External 20170714. *Figshare.* doi:10.6084/m9.figshare.5203936. Accessed date: mm/dd/yyyy. [The latest version is available at: http://tinyurl.com/DSMMguideEX]

Maturity Scale / Key Component	Level 1 - Ad Hoc Not Managed	Level 2 - Minimal Managed Limited	Level 3 - Intermediate Managed Defined, Partially Implemented	Level 4 - Advanced Managed Well-Defined, Fully Implemented	Level 5 - Optimal Level 4 + Measured , Controlled , Audit
Preservability	*The state of being preservable*				
Accessibility	*The state of being publicly searchable and accessible*				
Usability	*The state of data product being easy to understand and use*				
Production Sustainability	*The state of data production being sustainable and extendable*				
Data Quality Assurance	*The state of data product quality being assured/screened*				
Data Quality Control /Monitoring	*The state of data product quality being controlled and monitored*				
Data Quality Assessment	*The state of data product quality being assessed*				
Transparency /Traceability	*The state of being transparent, trackable, and traceable*				
Data Integrity	*The state of data integrity being verifiable*				

Figure 1: Diagram of DSMM maturity scale structure and key components.

2. Getting started

2.1. Assumptions

i. Datasets are digital environmental data products that are publicly available online.
ii. Evaluators who use this template have a basic knowledge of or are able to obtain information about conventions or standards relevant to practices examined in each key component in the community that datasets are produced for or/and provided to.

2.2. Steps for carrying out a self-evaluation of dataset stewardship maturity

i. Download the latest DSMM template file from http://tinyurl.com/DSMMtemplate.

ii. Rename the template file, using the following file naming convention:

<DatasetShortName>_MM-Stew_<vxxrxx>_<yyyymmdd>

The <…> denotes a field to be defined and filled in by evaluators based on information pertaining to the dataset.

<DatasetShortName> is a short name that is descriptive of the data product which, preferably, is 30 or less characters of letters, numbers, hyphen(s) and/or underscore(s) without any space or special characters. This short dataset name could include organization(s), data product abbreviation, and product type or/and version.

If applicable, it is recommended to use standard or abbreviated variable names such as those from Climate Convention (CF, e.g., http://cfconventions.org/Data/cf-standard-names/28/build/cf-standard-name-table.html) or Obs4MIPs (Observations for Model Intercomparisons, e.g., http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.pdf). For example, PMSIC for Passive Microwave Sea Ice Concentration. One may wish to include an abbreviation for institution(s) or program(s) such as EPA, USGS, NOAA NASA, or S-NPP, preferably using community standard-based keywords such as Global Change Master Directory (GCMD) keywords (http://gcmd.nasa.gov/learn/keywords.html), using hyphen if more than one primary institutions, for example, NOAA-NSIDC. It is recommended to consult, if it is possible, with product POC (Point-Of-Contact), scientific and/or data stewards for your choice of the short name.

<vxxrxx> is version and revision number of the current maturity assessment. For example, v01r00 will be used for the first baseline version. Version number only changes when maturity ratings are modified. Changing revision number reflects other modifications to the assessment document, including those to justifications. Both scoreboard and rating diagrams, if included in the document, will display maturity levels of the latest version and will not change with different revisions.

<yyyymmdd> is year, month, and day of the current maturity assessment version. For example, 20160408 for April 18, 2016.

Therefore, the file name:

NOAA-NSIDC_PMSIC_CDR-v2_MM-Stew_v02r01_20150623.pdf

will denote that it is the document containing the version v02r01 (i.e., version 2 with revision 1) stewardship maturity assessment results as of June 23, 2015 for the version 2 NOAA/NSIDC Sea Ice Concentration Climate Data Record (Appendix A).

iii. Go over the whole template file and read the disclaimer carefully before using the template.

iv. Enter dataset and relevant Point-Of-Contacts (POCs) information in the DSMM metadata section (Figure 2).

v. Read the content of DSMM in the template, if in question,

- Review high-level background information on DSMM at http://tinyurl.com/DSMMintro.

- Review the scope and rationale for each key component of the matrix including examples of community-accepted best practices and standards provided in Peng et al. (2015) at http://tinyurl.com/DSMMpaper.

vi. Go through each key component in the template, identify the stewardship practices applied to the dataset, and document your rating and justifications, including any relevant URLs. Add comment if necessary. Use dark brown text to denote uncertainty about the practice or the fact that additional information or research is needed.

- All stewardship practices applied to the dataset, not just ones by the repository/archive, should be captured.
- It is recommended to identify the community in the justification section that stewardship practices and standards are pertained to, based on the categories of communities defined in the template or defined in your own words.
- It is recommended to adopt community-accepted naming conventions for variable, sensor, or organizations, ... etc., such as these mentioned above in Section 2.2. under item ii.

vii. Obtain any additional information if necessary from product website(s), in the literature, and/or from POCs of dataset, archive, and access.

viii. Review the results and fill in the matrix cells with the defined color scheme (provided in Table I)

ix. Capture the assessment results in the DSMM metadata section

- One can wait to do the last two steps until after the assessment results are baselined.
- For simplicity, one can remove any other sections of the self-assessment template except for the maturity metadata section and ratings and justification section (see Figure 2) before the review process. The URLs to examples of assessment documents for some NCEI core datasets are provided in Appendix A-F (NOAA e-mail login required).

3. Assessment status levels

Very often, it is necessary to describe the current status of assessment to other team members or management. It would be good to have a consistent way to do so. To indicate the current status of the assessment, the assessment levels are recommended to progress in the following order:

- Not yet assessed,
- Preliminary assessment,
- Initial Assessment Draft (complete assessment but before the first (team) review);
- Revised Assessment Draft,
- Final Assessment Draft,
- Baselined.

The number of necessary drafts and their reviews before the first baseline varies on individual datasets, largely depending on how readily information about their stewardship practices is accessible to the evaluators.

It is recommended to version a draft of the assessment document as v00rxx, i.e.,

<DatasetShortName>_MM-Stew_<v00rxx>_<yyyymmdd>

Maturity Scale	Definition	Color Code	R	G	B	Color
Level 1	Ad Hoc/Unknown: Not Managed	Lighter Green	229	244	224	
Level 2	Minimal: Managed Limited	Light Green	203	234	192	
Level 3	Intermediate: Managed, defined, partially implemented	Green	176	223	161	
Level 4	Advanced: Managed, Well-defined, fully implemented	Dark Green	85	168	57	
Level 5	Optimal: Level 4 +, measured, controlled, audit	Darker Green	56	112	38	

Table I: Scientific Data Stewardship Maturity Matrix Scale Definition and RGB Color Scheme

4. Review and baseline process

The review is recommended to be done by the assessment team. For each data product, the assessment team may include a DSMM Subject Matter Expert (SME), data product SME or data manager, archive and access specialists assigned for the data product. (It was revealed from the previous pilot NCEI DSMM use case study that it is very beneficial to carry out assessment and review by a team – consisting of members from the Integrated Product Team, Data Steward, and the DSMM SME. Within the NCEI, an Integrated Product Team usually consists of a data product POC/SME, archive specialist, and access specialist.)

After the assessment draft review is finished and all the necessary revisions are made, the assessment ratings and justification will be baselined. Optionally, the stewardship scoreboard and rating diagram may be generated, utilizing the DSMM Graphics Generator (PC version, freely available for download at http://tinyurl.com/DSMMgraphics), and included in the baselined assessment document.

The first baseline version is normally versioned as v01r00. The pdf version of this document can then be used to demonstrate the stewardship maturity of the product. It is also a valuable resource for creating content-rich data quality metadata. Figure 2 shows an example of the layout of the assessment document using the assessment results for the NCEI Global Historical Climatology Network (GHCN)-Monthly version 3 data product. (The stewardship maturity assessment results and lessons learned from the use case study of NCEI GHCN-Monthly v3 (Peng et al., 2016) can be found at: http://tinyurl.com/DSMM-GHCN-Mv3)

Workflows and tools have been developed by the OneStop Project to standardize the evaluation, collection, presentation, and integration of assessments and to automate the generation of consistent assessment documents as data stewardship maturity reports (DSMRs). The effort is underway to publish these DSMRs as citable product quality descriptive information documents by NOAA Central Library. Once finalized, they will be described in the

NOAA internal version of this guide. The current version can be accessed at: http://tinyurl.com/DSMMguide (NOAA e-mail account login required.)

5. Integration to collection-level metadata and NOAA catalog

Dataset maturity rating information should be integrated into the standard-based collection-level metadata records and made accessible to other search or catalog services. Workflows described below pertain to NOAA OneStop only but may be used as references for other entities.

Eventually, a revised NCEI metadata tool system will provide the capability of systematically collecting this information. Prior to the availability of the system, maturity rating information may be collected via a Google Form:

http://tinyurl.com/DSMMsubmNew (select "MM-Stew" maturity metadata tag)

This form is currently only accessible to those who are at or affiliated with NOAA.

For people who are outside of NOAA and have been registered at:

http://tinyurl.com/DSMMregister

may provide your assessment results via a Google Form at:

http://tinyurl.com/DSMMsubmET

(send an e-mail to Ge.Peng@noaa.gov to request access after you have registered.)

The data product stewardship maturity information has been integrated in the dataset collection-level metadata record by the NOAA OneStop metadata team.

For enhanced interoperability, it is recommended to use the following maturity metadata (MM) tags as identifiers in the metadata records. These MM tags systematically indicate different perspectives for potential linkage to different defined maturity assessment models when such a system becomes available in the future:

- MM-Scie for science maturity information
- MM-Prod for product maturity information
- MM-Stew for stewardship maturity information
- MM-Serv for service maturity information

The mapping to ISO 19115 collection-level metadata has been implementation – the implement requirements and best practices are defined by the NOAA OneStop metadata team in collaboration with the NOAA Metadata Working Group.

NCEI collection-level metadata records are reviewed and published to the Web Accessible Folder (WAF) for other metadata resources such as NOAA catalog to harvest.

6. Acknowledgement

Support is provided by the NOAA OneStop Project under Cooperative Agreement NA14NES432003. Feedback from Paul Lemieux, Anna Milan, Nancy Ritchey, and Valerie Toner are beneficial. We thank David Pezdirtz for input on documentation versioning best practices.

7. Reference

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**. doi:10.2481/dsj.14-049. [Available online at: http://tinyurl.com/DSMMpaper]

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. [Available online at: http://tinyurl.com/DSMM-GHCN-Mv3]

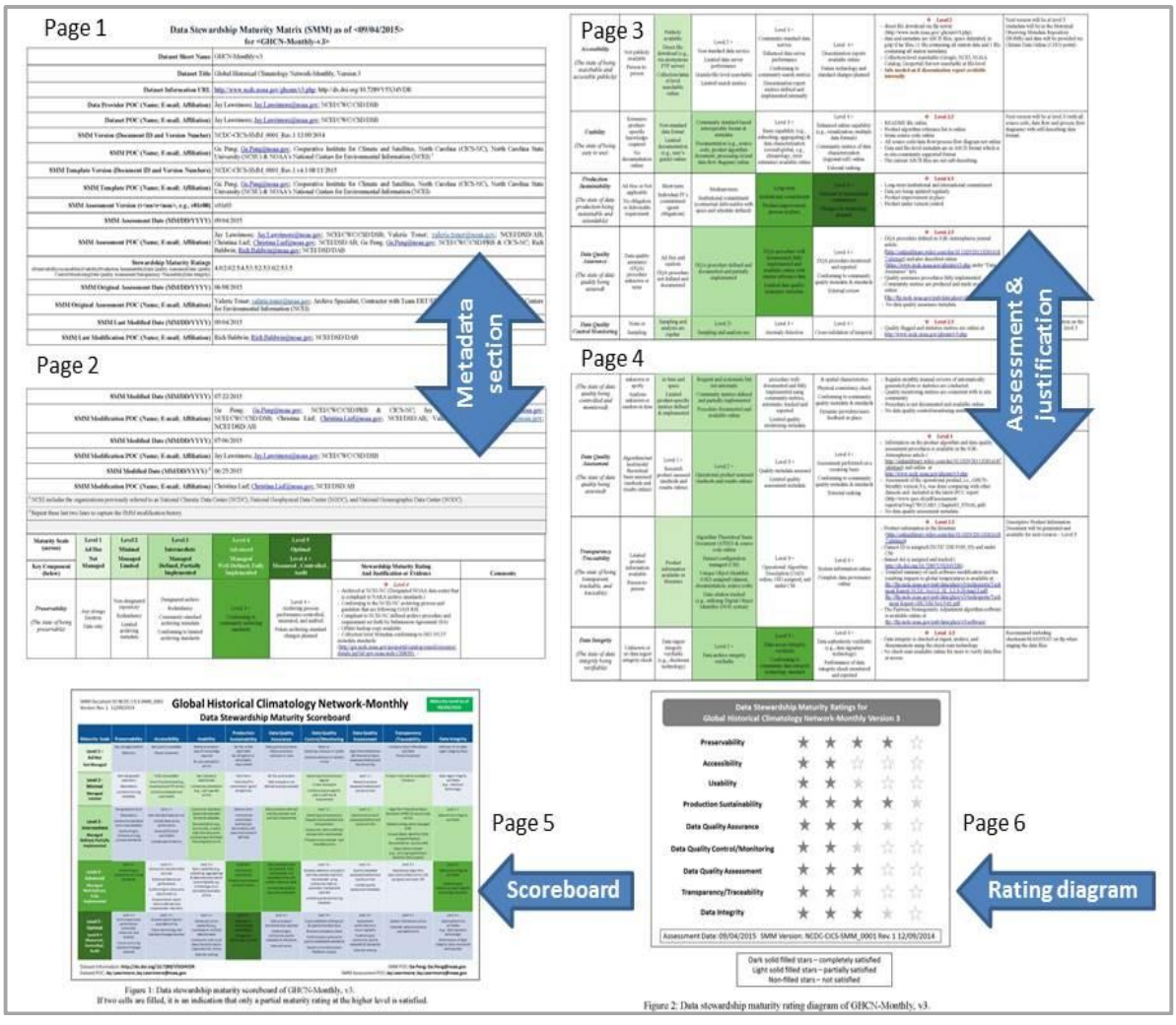

Figure 2: Page view of the stewardship maturity assessment document (6 pages in total) for NCEI GHCN-Monthly v3 data product to show the general layout of the document. It is not intended to display the content of the document, which can be viewed in Appendix B.

Appendix A

For those at or affiliated with NOAA, the latest PMSIC CDR MM-Stew document can be accessed at: http://tinyurl.com/MM-Stew-PMSIC-CDR.

Appendix B

For those at or affiliated with NOAA, the latest GHCN-M MM-Stew document can be accessed at: http://tinyurl.com/MM-Stew-GHCN-Mv3 (The maturity ratings and lessons learned from the DSMM GHCN-Monthly use case study can be found in Peng et al. (2016) and accessed at: http://tinyurl.com/DSMM-GHCN-Mv3)

Stewardship Maturity Matrix (SMM) as of <09/04/2015>
for <GHCN-M-v3>

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_0001_Rev.1 12/09/2014
SMM POC (Name; E-mail; Affiliation)	Ge Peng; Ge.Peng@noaa.gov; Cooperative Institute for Climate and Satellites, North Carolina (CICS-NC), North Carolina State University (NCSU) & NOAA's National Centers for Environmental Information (NCEI) [1]
SMM Template Version (Document ID and Version Numbers)	NCDC-CICS-SMM_0001_Rev.1 v4.0 06/23/2015
SMM Template POC (Name; E-mail; Affiliation)	Ge Peng; Ge.Peng@noaa.gov; Cooperative Institute for Climate and Satellites, North Carolina (CICS-NC), North Carolina State University (NCSU) & NOAA's National Centers for Environmental Information (NCEI)
SMM Assessment Version (v<nn>r<mm>, e.g., v01r00)	v01r03
SMM Assessment Date (MM/DD/YYYY)	09/04/2015
SMM Assessment POC (Name; E-mail; Affiliation)	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
Stewardship Maturity Ratings (kc1/kc2/kc3/kc4/kc5/kc6/kc7/kc8/kc9)	4.0/2.0/2.5/4.5/3.5/2.5/3.0/2.5/3.5
SMM Original Assessment Date (MM/DD/YYYY)	06/08/2015
SMM Original Assessment POC (Name; E-mail; Affiliation)	Valerie Toner; valerie.toner@noaa.gov; Archive Specialist, Contractor with Team ERT/STG, an affiliate of NOAA's National Centers for Environmental Information (NCEI)
SMM Last Modified Date (MM/DD/YYYY)	09/04/2015
SMM Last Modification POC (Name; E-mail; Affiliation)	Rich Baldwin; Rich.Baldwin@noaa.gov; NCEI/DSD/DAB
SMM Modified Date (MM/DD/YYYY)	07/22/2015

SMM Modification POC (Name; E-mail; Affiliation)	Ge Peng; Ge.Peng@noaa.gov; NCEI/CWC/CSD/PRB & CICS-NC; Jay Lawrimore; Jay.Lawrimore@noaa.gov; NCEI/CWC/CSD/DSB; Christina Lief; Christina.Lief@noaa.gov; NCEI/DSD/AB; Valerie Toner; valerie.toner@noaa.gov; NCEI/DSD/AB
SMM Modified Date (MM/DD/YYYY)	07/06/2015
SMM Modification POC (Name; E-mail; Affiliation)	Jay Lawrimore; Jay.Lawrimore@noaa.gov; NCEI/CWC/CSD/DSB
SMM Modified Date (MM/DD/YYYY) [2]	06/25/2015
SMM Modification POC (Name; E-mail; Affiliation)	Christina Lief; Christina.Lief@noaa.gov; NCEI/DSD/AB

[1] NCEI includes the organizations previously referred to as National Climatic Data Center (NCDC), National Geophysical Data Center (NGDC), and National Oceanographic Data Center (NODC).

[2] Repeat these last two lines to capture the SMM modification history

Maturity Scale (across) / **Key Component (below)**	**Level 1 Ad Hoc Not Managed**	**Level 2 Minimal Managed Limited**	**Level 3 Intermediate Managed Defined, Partially Implemented**	**Level 4 Advanced Managed Well-Defined, Fully Implemented**	**Level 5 Optimal Level 4 + Measured , Controlled , Audit**	**Stewardship Maturity Rating And Justification or Evidence**	**Comments**
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	❖ ***Level 4*** - Archived at NCEI-NC (Designated NOAA data center that is compliant to NARA archive standards.) - Conforming to the NCEI-NC archiving process and guideline that are following OAIS RM - Compliant to NCEI-NC defined archive procedure and requirement set forth by Submission Agreement (SA) - Offsite backup copy available - Collection level Metadata conforming to ISO 19115 metadata standards: (http://gis.ncdc.noaa.gov/geoportal/catalog/search/resource/details.jsp?id=gov.noaa.ncdc:C00839) ~	
Accessibility *(The state of being searchable and accessible publicly)*	Not publicly available Person-to-person	Publicly available Direct file download (e.g., via anonymous FTP server) Collection/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	❖ **Level 2** - direct file download via ftp server (http://www.ncdc.noaa.gov/ ghcnm/v3.php) - data and metadata are ASCII files, space delimited, in gzip'd tar files. (1 file containing all station data and 1 file containing all station metadata) - Collection-level searchable (Google, NCEI, NOAA Catalog, Geoportal) but not searchable at file-level - **Info needed on if dissemination report available internally**	Next version will be at level 3 (metadata will be in the Historical Observing Metadata Repository (HOMR) and data will be provided via Climate Data Online (CDO) portal)
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	❖ **Level 2.5** - README file online - Product algorithm reference list is online - Some source code online - All source code/data flow/process flow diagram not online - Data and file-level metadata are in ASCII format which is in situ community supported format - The current ASCII files are not self-describing	Next version will be at level 3 (with all source code, data flow and process flow diagrams) with self-describing data format.
Production Sustainability *(The state of data production being sustainable and extendable)*	Ad Hoc or Not applicable No 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 technology planned	❖ **Level 4.5** - Long-term institutional and international commitment - Data are being updated regularly - Product improvement in place - Product under version control	
Data Quality Assurance *(The state of data quality being assured)*	Data quality assurance (DQA) procedure unknown or none	Ad Hoc and random DQA 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	❖ **Level 3.5** - DQA procedure defined in JGR-Atmospheres journal article (http://onlinelibrary.wiley.com/doi/10.1029/2011JD016187/abstract) and also described online (https://www.ncdc.noaa.gov/ghcnm/v3.php under "Data Assurance" tab) - Quality assurance procedures fully implemented - Community metrics are produced and made available online (ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/products/) - No data quality assurance metadata	
Data Quality Control/Monitoring	None or Sampling	Sampling and analysis are regular	Level 2+ Sampling and analysis are	Level 3 + Anomaly detection	Level 4 + Cross-validation of temporal	❖ **Level 2.5** - Quality flagged and statistics metrics are online at: http://www.ncdc.noaa.gov/ghcnm/v3.php	For next version, documentation on the procedure(s) will be online – level 3

(The state of data quality being controlled and monitored)	unknown or spotty Analysis unknown or random in time	in time and space Limited product-specific metrics defined & implemented	frequent and systematic but not automatic Community metrics defined and partially implemented Procedure documented and available online	procedure well-documented and fully implemented using community metrics, automatic, tracked and reported Limited quality monitoring metadata	& spatial characteristics Physical consistency check Conforming to community quality metadata & standards Dynamic providers/users feedback in place	- Regular monthly manual reviews of automatically generated plots or statistics are conducted. - Quality monitoring metrics are consistent with in situ community - Procedure is not documented and available online - No data quality control/monitoring metadata	
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	❖ **Level 3** - Information on the product algorithm and data quality assessment procedures is available in the JGR-Atmospheres article (http://onlinelibrary.wiley.com/doi/10.1029/2011JD016187/abstract) and online at http://www.ncdc.noaa.gov/ghcnm/v3.php. - Assessment of the operational product, i.e., GHCN-Monthly version 3.x, was done comparing with other datasets and included in the latest IPCC report (http://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_Chapter02_FINAL.pdf) - No data quality assessment metadata	
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	❖ **Level 2.5** - Product information in the literature (http://onlinelibrary.wiley.com/doi/10.1029/2011JD016187/abstract) - Dataset ID is assigned (NCDC DSI 9100_03) and under CM - dataset doi is assigned and tracked (http://dx.doi.org/10.7289/V5X34VDR) - Detailed summary of each software modification and the resulting impacts to global temperatures is available at: ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical Report NCDC No112_02_3.2.0-29Aug12.pdf; ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical Report GHCNM No15-01.pdf - The Pairwise Homogeneity Adjustment algorithm software is available online at ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/software/	Descriptive Product Information Document will be generated and available for next version – Level 3
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	❖ **Level 3.5** - Data integrity is checked at ingest, archive, and dissemination using the check-sum technology - No check-sum available online for users to verify data files at access	Recommend including checksum/MANIFEST on ftp when staging the data files

SMM Document ID: NCDC-CICS-SMM_0001
Version: Rev. 1. 12/09/2014

Global Historical Climatology Network-Monthly

Data Stewardship Maturity Scoreboard

Maturity Level as of 09/04/2015

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://dx.doi.org/10.7289/V5X34VDR**
Dataset POC: **Jay Lawrimore; Jay.Lawrimore@noaa.gov**

SMM POC: **Ge Peng; Ge.Peng@noaa.gov**
SMM Assessment POC: **Jay Lawrimore; Jay.Lawrimore@noaa.gov**

Figure 1: Data stewardship maturity scoreboard of GHCN-M-v3.
If two cells are filled, it is an indication that only a partial maturity rating at the higher level is satisfied.

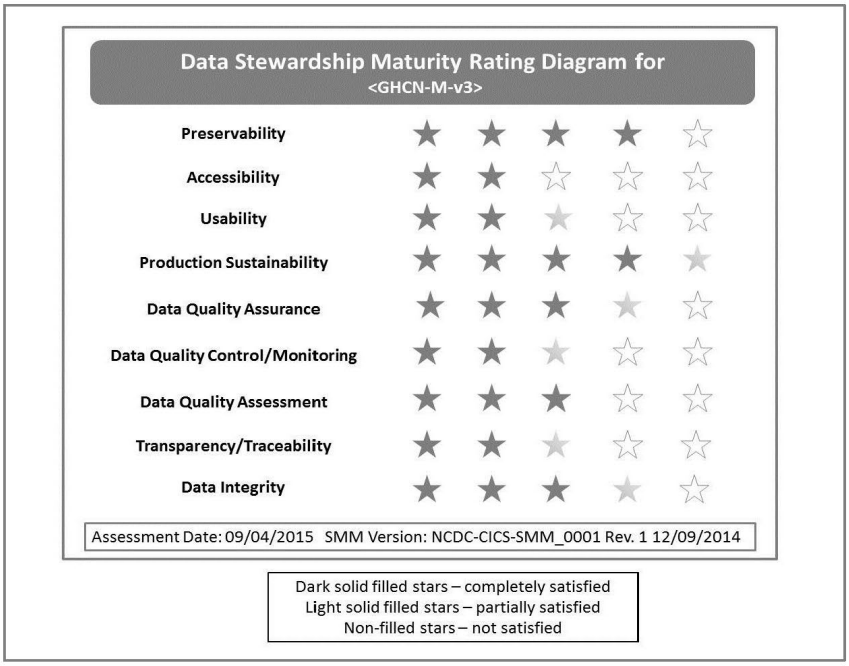

Figure 2: Data stewardship maturity rating diagram of GHCN-M-v3.

Appendix C

For those at or affiliated with NOAA, the latest NOAA/WDS ITRDB MM-Stew document can be accessed at: http://tinyurl.com/MM-Stew-ITRDB.

Appendix D

For those at or affiliated with NOAA, the latest NOAA Daily OISST CDR MM-Stew document can be accessed at: http://tinyurl.com/MM-Stew-DOISST-CDR.

Appendix E

For those at or affiliated with NOAA, the latest NOAA's NCEI DEMs MM-Stew document can be accessed at: http://tinyurl.com/MM-Stew-DEMs.

Appendix F

For those at or affiliated with NOAA, the latest NOAA's S-NPP VIIRS AF ARP MM-Stew document can be accessed at: http://tinyurl.com/MM-Stew-VIIRS-AF.