Towards Consistent and Citable Data Quality Descriptive Information for End-Users

Ge Peng[1,2], Nancy Ritchey[2], Anna Milan[2], Sonny Zinn[3], Kenneth S. Casey[2], David Neufeld[4], Paul Lemieux[3], Raisa Ionin[3], Robert Partee[3], Donald Collins[2], Jason Shapiro[3], Aaron Rosenberg[4], Thomas Jaensch[3], and Philip Jones[3]

[1] Cooperative Institute for Climate and Satellites – North Carolina (CICS–NC), NC State University, Asheville, NC, USA. Ge.Peng@noaa.gov; [2] NOAA's National Centers for Environmental Information (NCEI); [3] ERT, Inc./NCEI; [4] Cooperative Institute for Research in Environmental Sciences (CIRES)/NCEI

Introduction

Curating quality descriptive information and metadata for datasets is a necessary step toward meeting the transparency requirement and helping establish the credibility and trustworthiness of individual data products. This, however, has been a difficult challenge for the data management community due to the lack of a consistent assessment framework, process, and workflow. Furthermore, developing and implementing these require multi-domain knowledge and close cross-disciplinary collaboration.

This presentation will first introduce a data stewardship maturity matrix (DSMM) as a reference framework for assessing stewardship maturity of individual digital datasets. Using the DSMM as an example, this presentation will then demonstrate that it is possible to consistently and systematically curate and publish data quality descriptive information both as citable documents and within ISO metadata records for human and machine end-users. These consistent and citable documents and metadata records can be readily integrated into or linked by other systems and tools to be used, for example, for enhanced data discoverability and usability.

This presentation will also outline the progress made under the auspice of the NOAA OneStop project in the area of consistently curate, publish, integrate, and display data quality information.

DSMM: What, Scope, Who, How, …

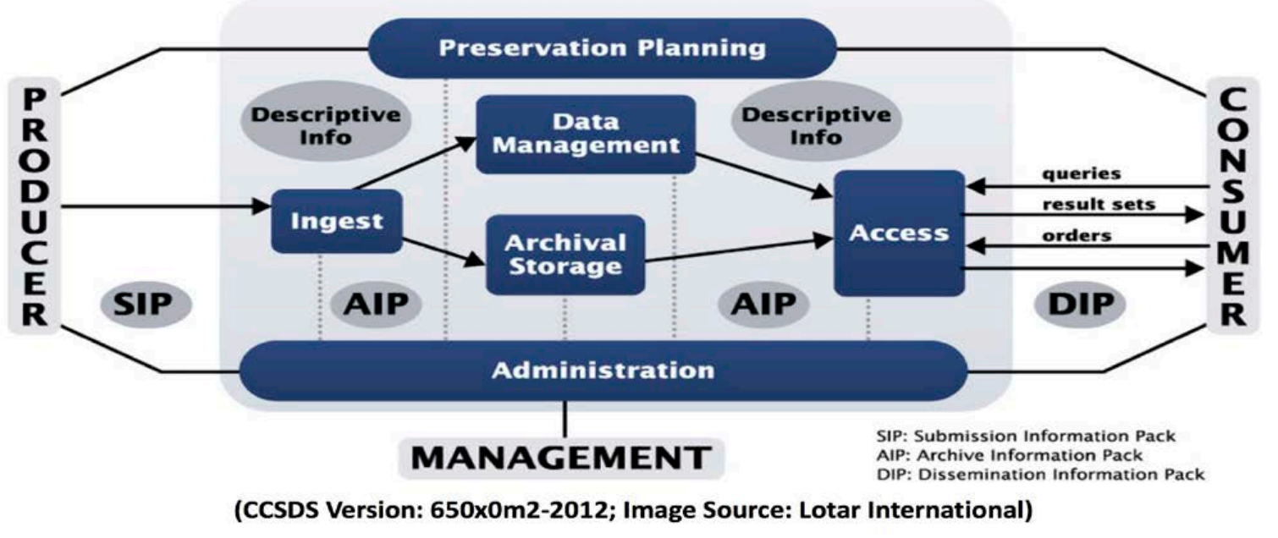
What Is the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix (DSMM)?

*A Unified Framework for Measuring **Stewardship Practices** Applied to Individual Digital Environmental Data Products That Are Publicly Available Online*

Leveraging Institutional Knowledge and Community Best Practices and Standards

The Scope of Stewardship Practices

Measureable practices associated with the functional entities of the Open Archival Information System (OAIS) (within the shaded box)



SIP: Submission Information Pack
AIP: Archive Information Pack
DIP: Dissemination Information Pack

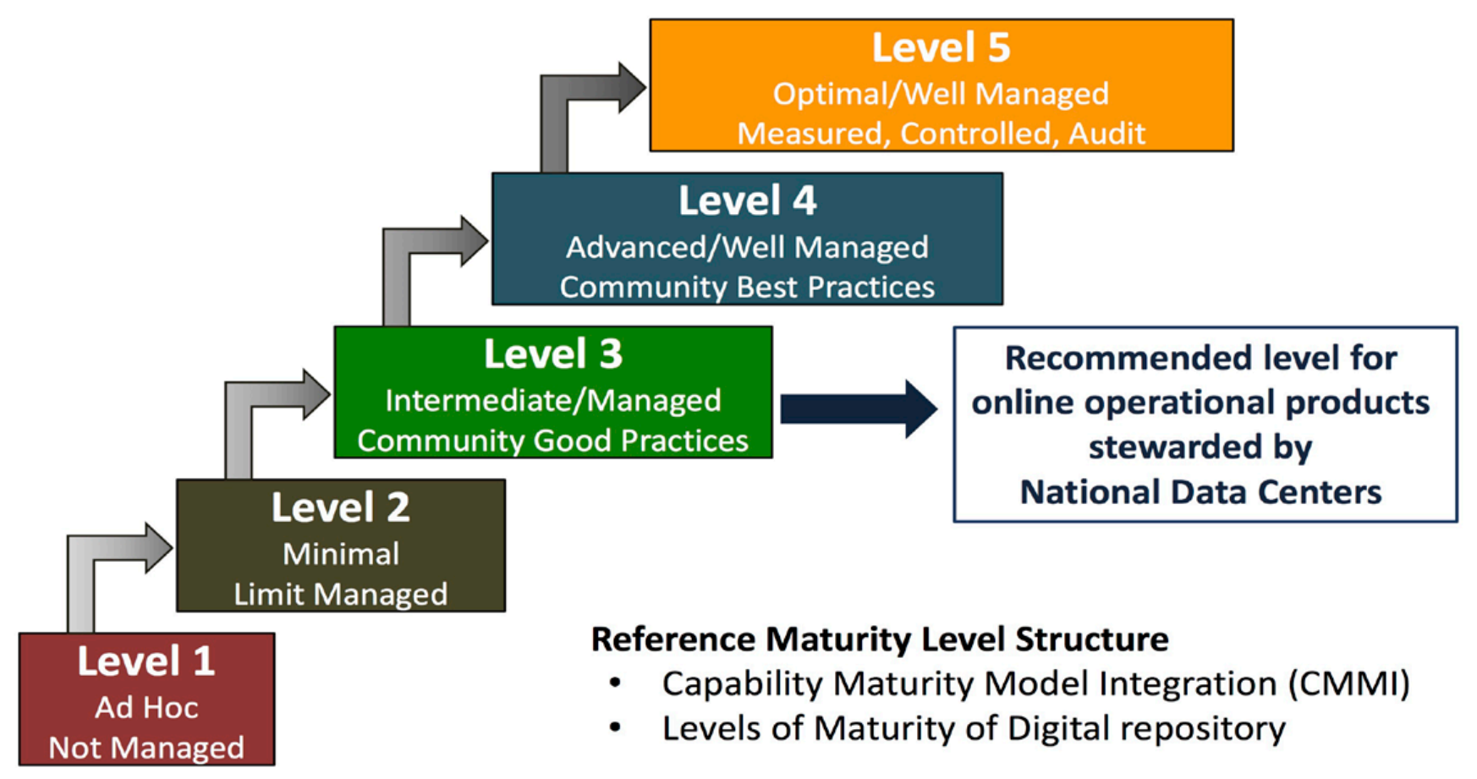
(CCSDS Version: 650x0m2-2012; Image Source: Lotar International)

DSMM Defines Measureable, Five-Level Progressive Practices

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 dataset being preservable*				
Accessibility	*The state of dataset 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 data product being transparent, trackable, and traceable*				
Data Integrity	*The state of data integrity being verifiable*				

(Data system integrity is also very important but not included in the matrix due to potential security risks to the system.)

DSMM Follows CMMI level Structure



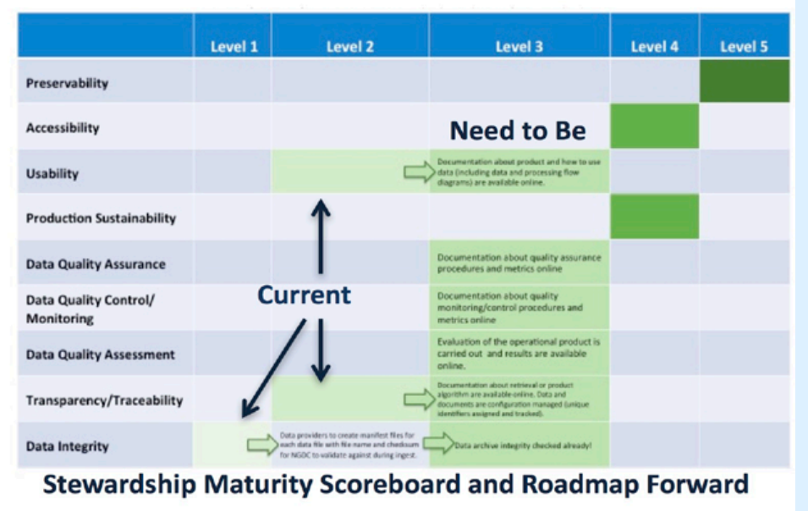
Reference Maturity Level Structure
- Capability Maturity Model Integration (CMMI)
- Levels of Maturity of Digital repository

Who Could Use The Matrix?

- **Data providers and scientific stewards**
  - ➢ to evaluate and improve the quality and usability of their products against community best practices
- **Modelers, decision-support system users, and scientists**
  - ➢ to improve their products and uncertainty estimates
  - ➢ to make investment and use decision
- **Data managers/stewards of data centers and repositories**
  - ➢ to validate their compliance or lack of to community accepted stewardship practice or standards
  - ➢ to assess the current state
  - ➢ to create a roadmap forward to improve or enhance its stewardship maturity of practices applied to a certain product or all its holdings
- **General data users**
  - ➢ to make an educated choice on selecting or utilizing a dataset

Ways to Utilize DSMM & Assessment Results

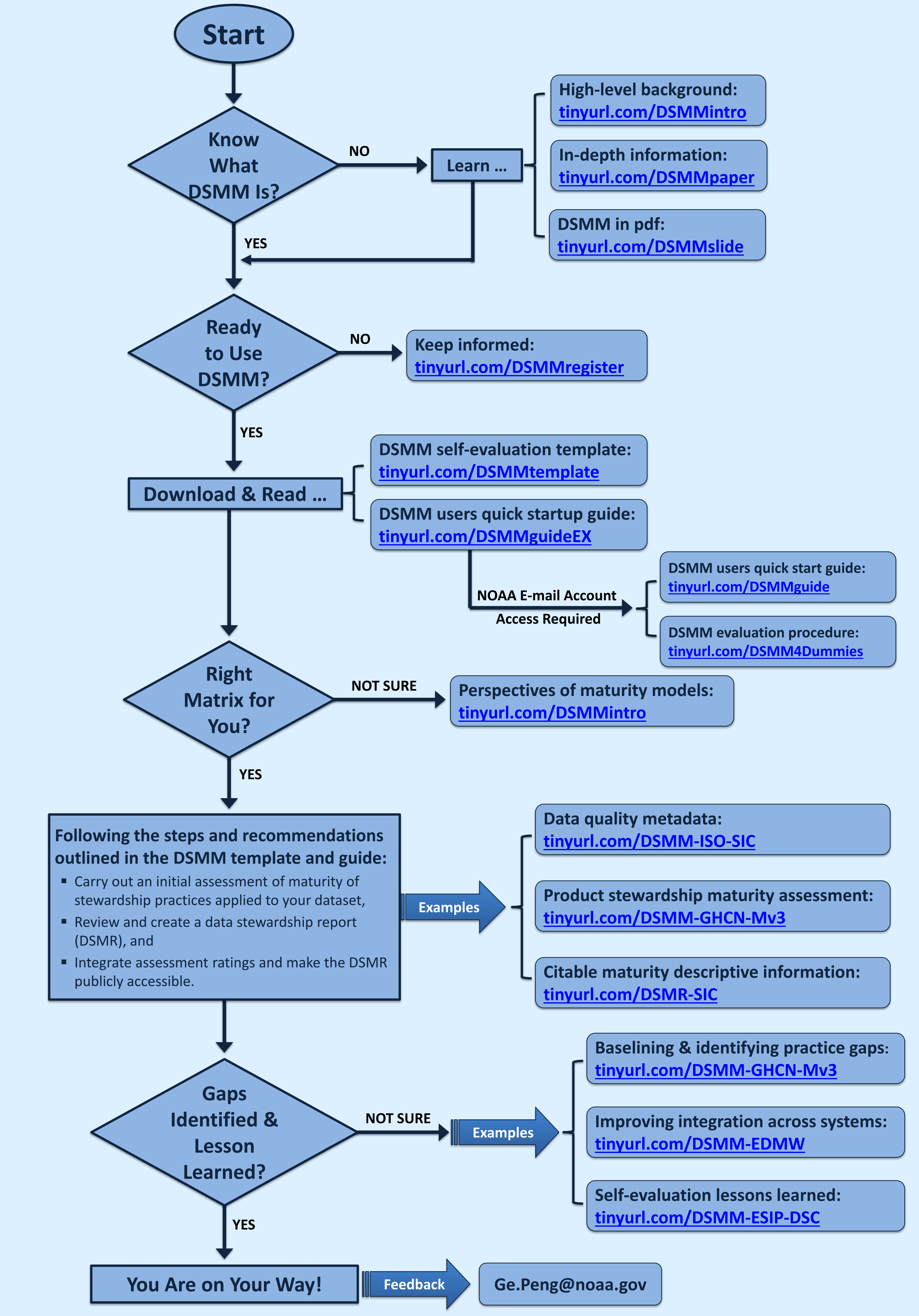
- **To know the current state of your dataset(s) – maturity assessment (stewardship maturity scoreboard)**
- **To know where you want or need to be – stewardship requirements**
- **To know how to get there – roadmap forward (informed, actionable steps)**


Level 1
Level 2
Level 3
Level 4
Level 5
Preservability
Accessibility
Usability
Production Sustainability
Data Quality Assurance
Data Quality Control/ Monitoring
Data Quality Assessment
Transparency/Traceability
Data Integrity
Need to Be
Current

Stewardship Maturity Scoreboard and Roadmap Forward

- **A reference model for stewardship planning and resource allocation – informed decision-making support**
- **A consolidate source and transparency for information about stewardship practices – assessment with detailed justifications**
- **Content-rich quality metadata – enhanced discoverability and usability**

Getting to Know and to Use the Data Stewardship Maturity Matrix (DSMM)



High-level background: tinyurl.com/DSMMintro

In-depth information: tinyurl.com/DSMMpaper

DSMM in pdf: tinyurl.com/DSMMslide

Keep informed: tinyurl.com/DSMMregister

DSMM self-evaluation template: tinyurl.com/DSMMtemplate

DSMM users quick startup guide: tinyurl.com/DSMMguideEX

DSMM users quick start guide: tinyurl.com/DSMMguide

DSMM evaluation procedure: tinyurl.com/DSMM4Dummies

Perspectives of maturity models: tinyurl.com/DSMMintro

Following the steps and recommendations outlined in the DSMM template and guide:
- Carry out an initial assessment of maturity of stewardship practices applied to your dataset,
- Review and create a data stewardship report (DSMR), and
- Integrate assessment ratings and make the DSMR publicly accessible.

Data quality metadata: tinyurl.com/DSMM-ISO-SIC

Product stewardship maturity assessment: tinyurl.com/DSMM-GHCN-Mv3

Citable maturity descriptive information: tinyurl.com/DSMR-SIC

Baselining & identifying practice gaps: tinyurl.com/DSMM-GHCN-Mv3

Improving integration across systems: tinyurl.com/DSMM-EDMW

Self-evaluation lessons learned: tinyurl.com/DSMM-ESIP-DSC

Ge.Peng@noaa.gov

Pathway to Application of DSMM

Communication and Use Case Studies

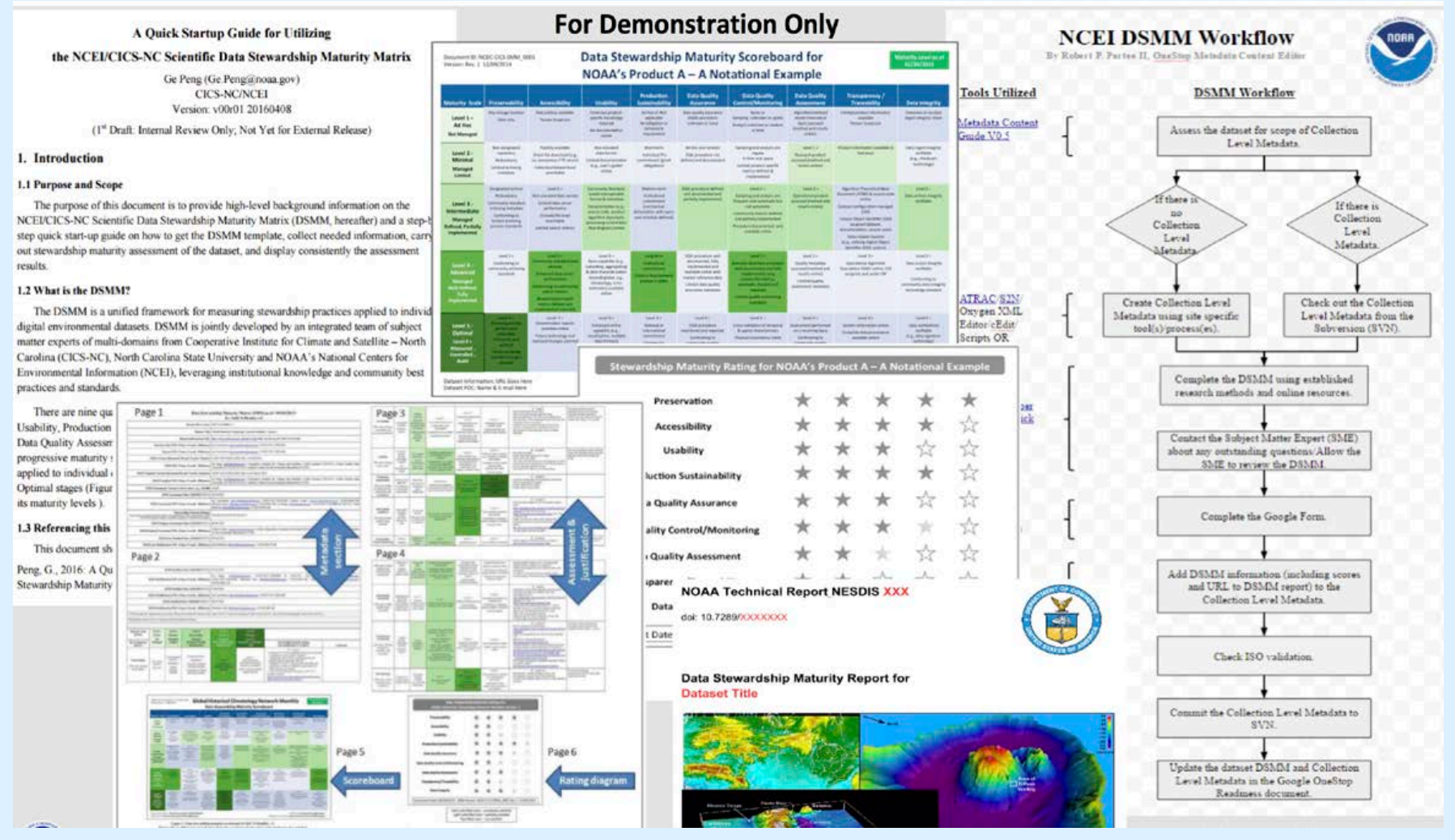
Selected NCEI Core Datasets

Data Type	Dataset	Status
Satellite – polar ocean	NOAA/NSIDC Sea Ice Concentration CDR	Baselined
GIS - regional	NCEI-CO Digital Elevation Models (DEM)	Revised assessment draft review
Station - in situ - land	GHCN-M	Baselined
Station - gridded - land	National Climate Division (nCliDiv)	Not yet started
Satellite – global ocean	Optimum Interpolation Sea Surface Temperature (OISST) CDR	Baselined
Physical Records - In Situ Monthly Summaries	Local Climatological Data	Initial assessment draft review
Paleo – global land	NOAA/WDS International Tree-Ring Data Bank (ITRDB)	Baselined

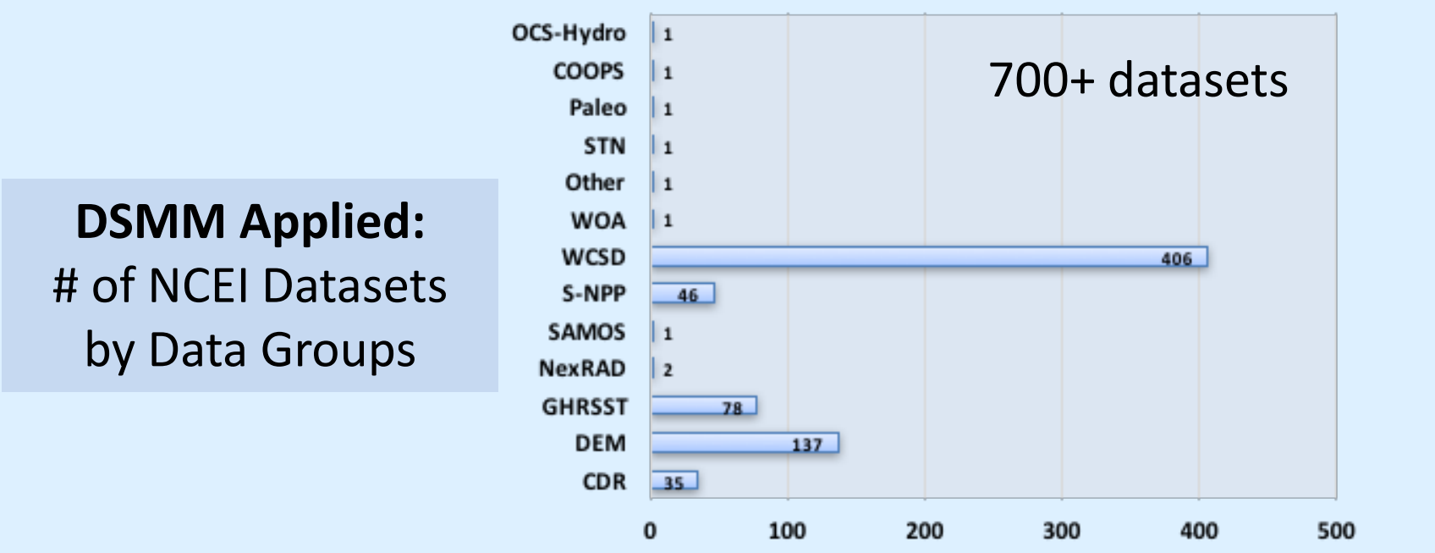
Selected ESIP Datasets

Data Type	Dataset	Status
Model Reanalysis	NCAR Global Climate Four-Dimensional Data Assimilation Hourly 40km Reanalysis	Baselined *
Ecological Data	DataOne Member Node SBC LTER (Long Term Ecological Research) Network	Revised assessment draft review
Long-tail Data	NSF ACADIS (Advanced Cooperative Arctic Data and Information Service)	Initial assessment draft
Socioeconomic Data	NASA Socioeconomic Data	Initial assessment draft
Paleo Data	Australia Borehole Data	Not yet assessed

Dataset Stewardship Maturity Evaluation: Guidance, Template, Training, and Tools


A Quick Startup Guide for Utilizing the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix
For Demonstration Only
Data Stewardship Maturity Scoreboard for NOAA's Product A – A Notational Example
Stewardship Maturity Rating for NOAA's Product A – A Notational Example
NOAA Technical Report NESDIS XXX
Data Stewardship Maturity Report for Dataset Title
NCEI DSMM Workflow
Scoreboard
Rating diagram

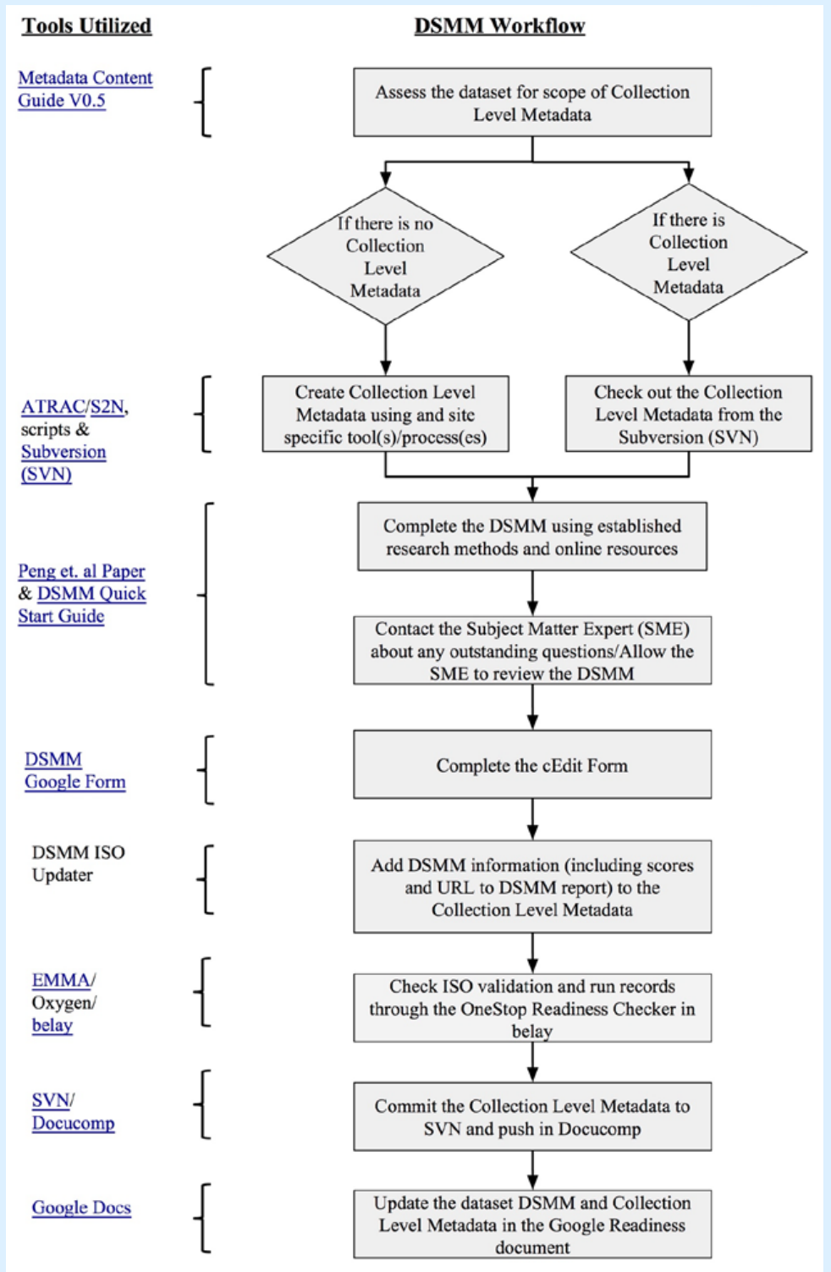
DSMM Applied: # of NCEI Datasets by Data Groups



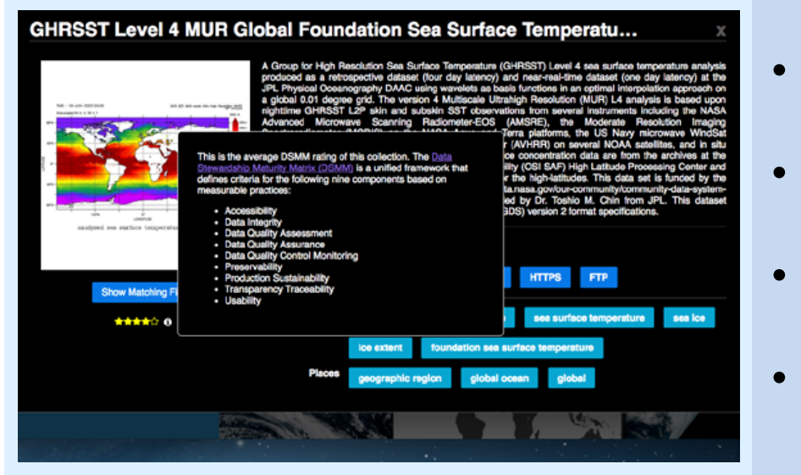
Consistent and Citable Data Quality Descriptive Information

Developing and Improving Evaluation, Curation, and Integration Workflows

DSMM Integration Workflow

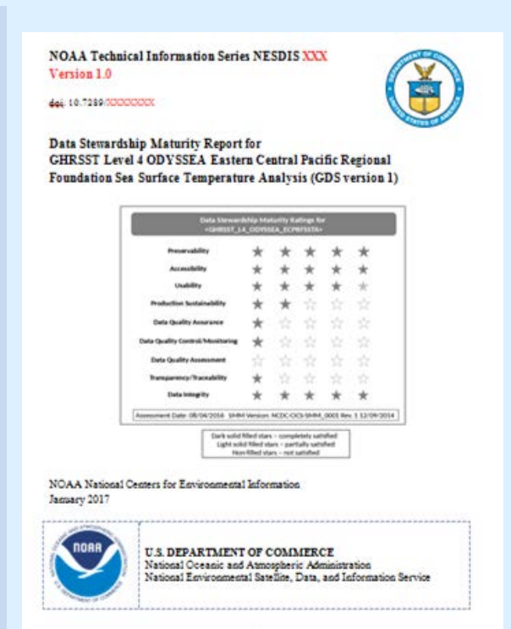

Tools Utilized
DSMM Workflow
Metadata Content Guide V0.5
Assess the dataset for scope of Collection Level Metadata
If there is no Collection Level Metadata
If there is Collection Level Metadata
ATRAC/S2N, scripts & Subversion (SVN)
Create Collection Level Metadata using site specific tool(s)/process(es)
Check out the Collection Level Metadata from the Subversion (SVN)
Peng et. al Paper & DSMM Quick Start Guide
Complete the DSMM using established research methods and online resources
Contact the Subject Matter Expert (SME) about any outstanding questions/Allow the SME to review the DSMM
DSMM Google Form
Complete the cEdit Form
DSMM ISO Updater
Add DSMM information (including scores and URL to DSMM report) to the Collection Level Metadata
EMMA/ Oxygen/ belay
Check ISO validation and run records through the OneStop Readiness Checker in belay
SVN/ Docucomp
Commit the Collection Level Metadata to SVN and push in Docucomp
Google Docs
Update the dataset DSMM and Collection Level Metadata in the Google Readiness document

Consistent Rating Displacement

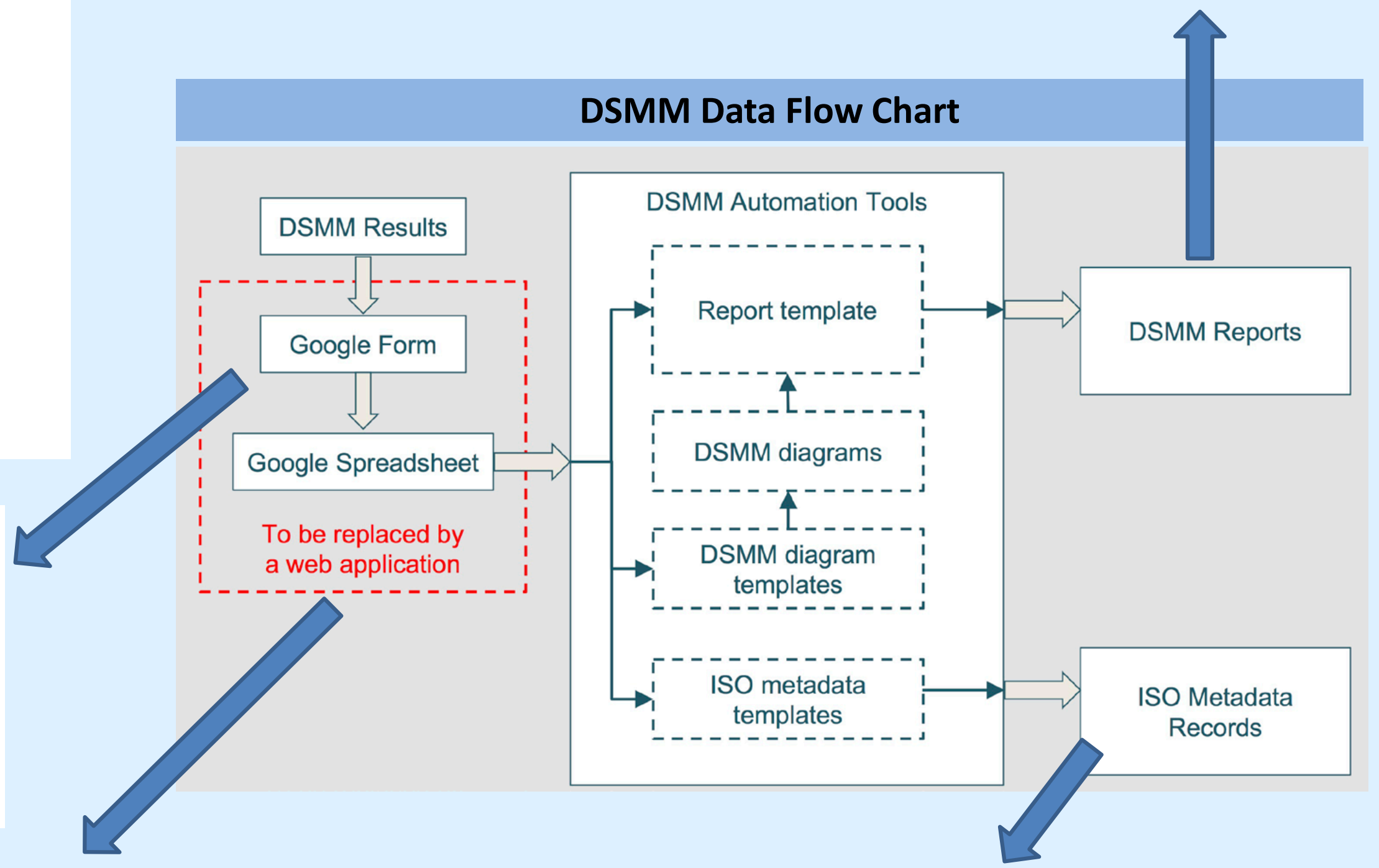

GHRSST Level 4 MUR Global Foundation Sea Surface Temperatu…
This is the average DSMM rating of this collection.
Accessibility
Data Integrity
Data Quality Assessment
Data Quality Assurance
Data Quality Control Monitoring
Preservability
Production Sustainability
Transparency Traceability
Usability
Show Matching
HTTPS
FTP
sea surface temperature
sea ice
ice extent
foundation sea surface temperature
Places
geographic region
global ocean
global

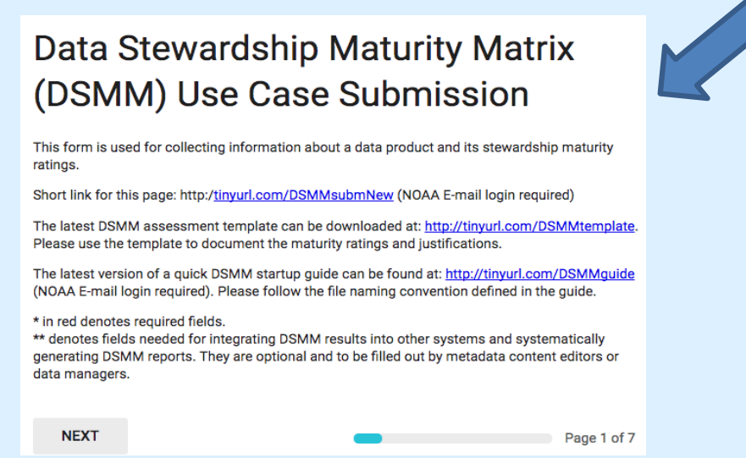
Data Stewardship Maturity Reports (DSMR)

- Data product quality descriptive information documents,
- Consistent document layout,
- Automated generation workflow,
- Unique persistent document identifier,
- To be published and archived by the NOAA Central Library.


NOAA Technical Information Series NESDIS XXX
Version 1.0
Data Stewardship Maturity Report for GHRSST Level 4 ODYSSEA Eastern Central Pacific Regional Foundation Sea Surface Temperature Analysis (GDS version 1)
NOAA National Centers for Environmental Information
January 2017
U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Environmental Satellite, Data, and Information Service

DSMM Data Flow Chart





Data Stewardship Maturity Matrix (DSMM) Use Case Submission

This form is used for collecting information about a data product and its stewardship maturity ratings.

Short link for this page: http:/tinyurl.com/DSMMsubmNew (NOAA E-mail login required)

The latest DSMM assessment template can be downloaded at: http://tinyurl.com/DSMMtemplate. Please use the template to document the maturity ratings and justifications.

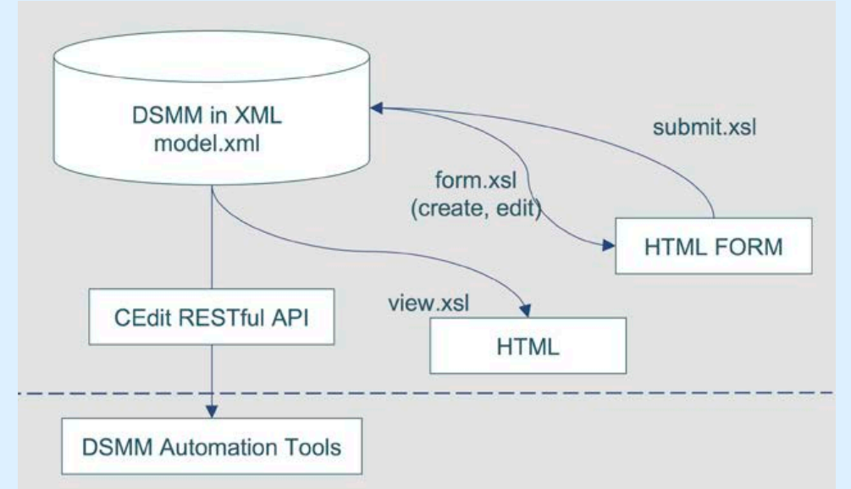
The latest version of a quick DSMM startup guide can be found at: http://tinyurl.com/DSMMguide (NOAA E-mail login required). Please follow the file naming convention defined in the guide.

* in red denotes required fields.
** denotes fields needed for integrating DSMM results into other systems and systematically generating DSMM reports. They are optional and to be filled out by metadata content editors or data managers.

NEXT

Page 1 of 7

Integration with CEdit



Consistent DSMM ISO 19115 Metadata Implementation

Conceptual Consistency

Template Content

Measure Name: Data Stewardship Maturity Assessment
Measure ID: MM-Stew
Measure Description: The Data Stewardship Maturity Matrix (DSMM) is a unified framework that defines criteria for each of nine components based on measurable practices, which can be used to apply a progressive, 6-level rating to an individual dataset, representing stewardship maturity stages rated as Not Assessed or Not Available (Level 0), adHoc (Level 1), minimum (Level 2), intermediate (Level 3), advanced (Level 4), and optimal (Level 5).
Evaluation Description: Data Stewardship Maturity Assessment was evaluated by the metadata content editor for the NOAA OneStop project using the Scientific Data Stewardship Maturity Assessment Model Template v4.0.
Procedure Reference: Peng, Ge. The Scientific Data Stewardship Maturity Assessment Model Template. 2015-06-23. doi:10.6084/m9.figshare.1211954

Conceptual Consistency

Results

Date of Measurement: 2016-06-22
Quantitative Result:
- Data Quality Assessment: minimal
- Accessibility: advanced
- Data Quality Control Monitoring: minimal
- Production Sustainability: advanced
- Data Integrity: intermediate
- Preservability: advanced
- Transparency Traceability: intermediate
- Usability: advanced
- Data Quality Assurance: advanced

Conformance Result
Explanation: Data Stewardship Maturity Assessment was (etc…)
Pass: (inapplicable)

Results

Reference: 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, NOAA National Centers for Environmental Information, doi: 10.7289/XXXXXX.

Acknowledgment

This work is supported by the NOAA OneStop Project. The ESIP DSMM use case study was sponsored by ESIP Data Stewardship Committee. The NCEI use case study was sponsored by NCEI Data Stewardship Division and Center for Weather and Climate.

NOAA National Centers for Environmental Information

Download the poster at https://ncics.org/dsmm

DataOne User Group Meeting, 24 July 2017