

Providing Rich and Structured Dataset Quality Information

Practical Application of a Data Stewardship Maturity Matrix

Ge Peng¹, Anna Milan², Nancy A. Ritchey², Robert P. Partee II³, Sonny Zinn³, Evan McQuinn⁴, Kenneth S. Casey², Paul Lemieux III³, Raisa Ionin³, Philip Jones³, Arianna Jakositz⁴, and Donald Collins²

¹ North Carolina State University, Cooperative Institute for Climate and Satellites - North Carolina (CICS-NC) at NOAA's National Centers for Environmental Information, Asheville, North Carolina 28801, USA;

² NOAA's National Centers for Environmental Information, USA; ³ Earth Resources Technology, Inc., USA; ⁴ Cooperative Institute for Research in Environmental Sciences (CIRES), USA

Why Dataset Quality Information Is Important?

- Data are an organizational asset
- Quality of datasets and associated information is fundamental for the quality of data services, ensuring trustworthiness of data holdings, and improving data management and stewardship

Why Accessible, Usable and Interoperable?

- U.S. laws and regulations require timely access to high-quality, readily usable, and interoperable federally funded data and information
- Structured machine- and human-readable dataset quality information helps ensure datasets are findable, accessible, interoperable AND reusable

What Is the DSMM?

- Consistent framework for assessing quantifiable stewardship practices
- Developed jointly by domain experts leveraging institutional knowledge and community best practices and standards
- Vetted through use case studies with diverse datasets managed by different organizations, in collaboration with NCEI Data Stewardship Division and ESIP Data Stewardship Committee

DSMM Evaluates Stewardship Maturity in Nine Key Components

- *Preservability*
- *Accessibility*
- *Usability*
- *Production Sustainability*
- *Data Quality Assurance*
- *Data Quality Control/Monitoring*
- *Data Quality Assessment*
- *Transparency/Traceability*
- *Data Integrity*

What Is the NOAA OneStop Project?

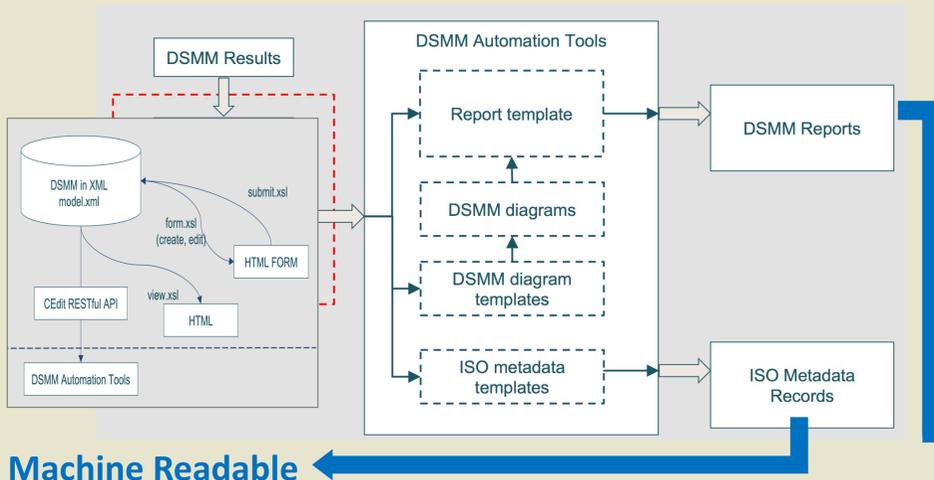
- <https://data.noaa.gov/onestop>
- Initiated in 2015
- Supports NOAA's efforts to improve discovery and access services for data
- Aims to enable broader use and reuse of NOAA data in commercial and scientific applications
- Leverages existing access technologies
- Infuses specific innovations
- Implements services in open data framework at massive scale

Data Stewardship Maturity Questionnaire (DSMQ)

Streamlined Assessment Process

Maturity Scale	Accessibility
Level 1 – Ad Hoc Not Managed	Not publicly available Person-to-person
Level 2 – Minimal Managed Limited	Publicly available Direct file download (e.g., via anonymous FTP server) Collection/dataset level searchable
Level 3 – Intermediate Managed Defined, Partially Implemented	Level 2 + Non-standard data service Limited data server performance Granule/file level searchable Limited search metrics
Level 4 – Advanced Managed Well-Defined, Fully Implemented	Level 3 + Community-standard data services Enhanced data server performance Conforming to community search metrics Dissemination report metrics defined and implemented internally
Level 5 – Optimal Level 4 + Measured, Controlled, Audit	Level 4 + Dissemination reports available online Future technology and standard changes planned

DSMM Data Flow Chart



Machine Readable

ISO Quality Metadata

<p>Conceptual Consistency</p> <p>Measure Name: Data Stewardship Maturity Assessment</p> <p>Measure ID: MM-Stew</p> <p>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).</p> <p>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.</p> <p>Procedure Reference: Peng, Ge. The Scientific Data Stewardship Maturity Assessment Model Template. 2015-06-23. doi:10.6084/m9.figshare.1211954</p>	<p>Conceptual Consistency</p> <p>Date of Measurement: 2016-06-22</p> <p>Quantitative Result:</p> <ul style="list-style-type: none"> • 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 <p>Conformance Result</p> <p>Explanation: Data Stewardship Maturity Assessment was (etc...)</p> <p>Pass: (inapplicable)</p> <p>Reference: Ionin, R., G. Peng, and K. Saha (2016), Data stewardship maturity report for GHRSS 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.</p>
--	--

Human Readable

Data Stewardship Maturity Reports (DSMR)

- Data product quality descriptive information documents
- Content-rich
- Consistent document layout
- Automated generation workflow
- Unique persistent document identifier (pending)

Take Away Messages

- Maturity assessment models can be used to measure and present quality ratings of individual datasets.
- Structured, evidence-based, content-rich, machine- and human-readable dataset quality information can be curated systematically.
- Dataset quality information helps ensure FAIR datasets—findable, accessible, interoperable, and reusable!

More Info and Resources

- ncics.org/dsmm
- gpeng@ncsu.edu
- twitter.com/DrPengAtAVL

Acknowledgement

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. Communication assistance provided by CICS-NC.

