



# Practical Application of the Data Stewardship Maturity Model for NOAA's *OneStop* Project

Nancy A. Ritchey<sup>1</sup>, Ge Peng<sup>2</sup>, Philip Jones<sup>3</sup>, Anna Milan<sup>1</sup>, Paul Lemieux<sup>4</sup>, Robert Partee<sup>4</sup>,  
Raisa Ionin<sup>4</sup>, Kenneth S. Casey<sup>1</sup>

<sup>1</sup>NOAA's National Centers for Environmental Information, <sup>2</sup>Cooperative Institute for Climate Studies (CICS-NC), <sup>3</sup>STG, Inc., <sup>4</sup>ERT

15 December 2016 15:25-15:40 Moscone West - 2000

NOAA Satellite and Information Service | National Centers for Environmental Information





# Outline

- What is *OneStop*?
- What is the Data Stewardship Maturity Model (DSMM)?
- DSMM relationship to Data Quality
- DSMM assessment and references
- Implementation of DSMM in ISO metadata
- Lessons learned



# NOAA's *OneStop Project*

- Led by NCEI
- Purpose to improve discovery of, access to, and increased usability of NOAA data, and is a NESDIS contribution to NOAA's open data and data stewardship efforts.
- A key aspect is improving all layers of data discovery and access framework including the supporting catalog services, data visualization and subsetting services, metadata, data formats and standards.



# DSMM - Data Stewardship Maturity Matrix

*A Unified Framework for  
Measuring **Stewardship Practices** Applied to  
Individual Digital Earth Sciences Data Products*

<http://tinyurl.com/DSMMintro>



# DSMM - A Consistent Framework for

- Assessing individual dataset **stewardship quality** - the current state of how datasets are documented, preserved, stewarded, and made accessible publicly.
- Capturing assessment results consistently for enhanced transparency.
- Allowing better data quality information integration and content-based search and discovery of NOAA data.

# DSMM Defines Measurable, Five-Level Progressive Practices in Nine Quasi-Independent Key Components

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	<i>The state of being preservable</i>				
Accessibility	<i>The state of being publicly searchable and accessible</i>				
Usability	<i>The state of data product being easy to understand and use</i>				
Production Sustainability	<i>The state of data production being sustainable and extendable</i>				
Data Quality Assurance	<i>The state of data product quality being assured/screened</i>				
Data Quality Control /Monitoring	<i>The state of data product quality being controlled and monitored</i>				
Data Quality Assessment	<i>The state of data product quality being assessed</i>				
Transparency /Traceability	<i>The state of being transparent, trackable, and traceable</i>				
Data Integrity	<i>The state of data integrity being verifiable</i>				

# DSMM Defines Measurable, Five-Level Progressive Practices in Nine Quasi-Independent Key Components

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	<i>The state of being preservable</i>				
Accessibility	<i>The state of being publicly searchable and accessible</i>				
Usability	<i>The state of data product being easy to understand and use</i>				
Production Sustainability	<i>The state of data production being sustainable and extendable</i>				
Data Quality Assurance	<i>The state of data product quality being assured/screened</i>				
Data Quality Control /Monitoring	<i>The state of data product quality being controlled and monitored</i>				
Data Quality Assessment	<i>The state of data product quality being assessed</i>				
Transparency /Traceability	<i>The state of being transparent, trackable, and traceable</i>				
Data Integrity	<i>The state of data integrity being verifiable</i>				



# Guidance, Templates, Tools

## 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: v00r01 20160408

(1<sup>st</sup> Draft: Internal Review Only; Not Yet for External Release)

### 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, hereafter) 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.

#### 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 Usability, Production Sustainability Data Quality Assessment, Transparency progressive maturity scale is defined applied to individual datasets, represent Optimal stages (Figure 1, also see its maturity levels).

#### 1.3 Referencing this document

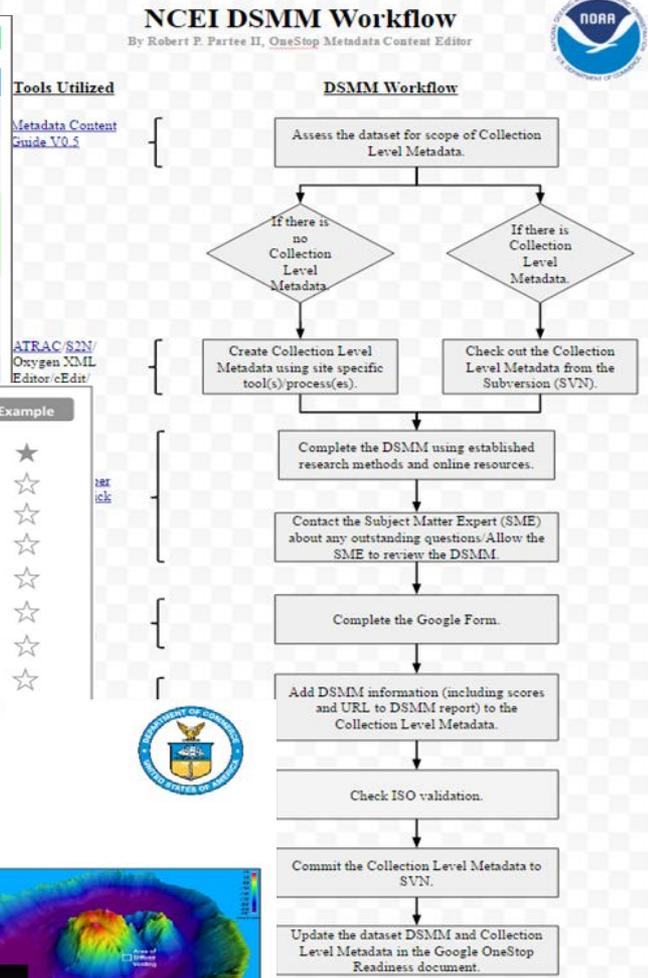
This document should be referenced as: Peng, G., 2016: A Quick Startup Guide for Utilizing the NCEI/CICS-NC Scientific Data Stewardship Maturity Matrix. Ver.

Document ID: NCEI-CICS-SMM\_0001  
Version: Rev. 1.12/09/2014

### Data Stewardship Maturity Scoreboard for NOAA's Product A – A Notational Example

Maturity Level as of 01/30/2015

Maturity Scale	Preservability	Accessibility	Usability	Production Sustainability	Data Quality Assurance	Data Quality Control/Monitoring	Data Quality Assessment	Transparency / Traceability	Data Integrity
<b>Level 1 – Ad Hoc / Not Managed</b>	Any change requires manual intervention	Not publicly available. Access to dataset is restricted	Dataset access is limited to specific knowledge required. No documentation is present	Dataset is not publicly available. Access to data is restricted. No documentation is present	Dataset quality is not assured. No quality assurance or control is in place	None or minimal sampling, collection or quality assurance activities are in place	Aggregating and analyzing data is manual and error-prone	Dataset information is not available. No metadata is present	Dataset integrity is not assured. No backup or recovery plan is in place
<b>Level 2 – Minimal / Managed / Limited</b>	Non-authorized users are not allowed to access the dataset	Publicly available. Access to dataset is restricted	Non-authorized users are not allowed to access the dataset. Limited documentation is present	Short-term dataset is available. Limited documentation is present	All data and metadata are documented. Data provenance is documented and maintained	Sampling and analysis are manual and error-prone. No quality assurance or control is in place	Reasonable effort is made to ensure data quality. Limited product specific quality assurance is in place	Product information is available in metadata	Dataset integrity is not assured. No backup or recovery plan is in place
<b>Level 3 – Intermediate / Managed / Defined / Partially Implemented</b>	Dataset access is restricted. Non-authorized users are not allowed to access the dataset	Publicly available. Access to dataset is restricted	Non-authorized users are not allowed to access the dataset. Limited documentation is present	Short-term dataset is available. Limited documentation is present	All data and metadata are documented. Data provenance is documented and maintained	Sampling and analysis are manual and error-prone. No quality assurance or control is in place	Reasonable effort is made to ensure data quality. Limited product specific quality assurance is in place	Product information is available in metadata	Dataset integrity is not assured. No backup or recovery plan is in place
<b>Level 4 – Fully Implemented / Managed / Fully Implemented</b>	Dataset access is restricted. Non-authorized users are not allowed to access the dataset	Publicly available. Access to dataset is restricted	Non-authorized users are not allowed to access the dataset. Limited documentation is present	Short-term dataset is available. Limited documentation is present	All data and metadata are documented. Data provenance is documented and maintained	Sampling and analysis are manual and error-prone. No quality assurance or control is in place	Reasonable effort is made to ensure data quality. Limited product specific quality assurance is in place	Product information is available in metadata	Dataset integrity is not assured. No backup or recovery plan is in place
<b>Level 5 – Optimal / Managed / Fully Implemented</b>	Dataset access is restricted. Non-authorized users are not allowed to access the dataset	Publicly available. Access to dataset is restricted	Non-authorized users are not allowed to access the dataset. Limited documentation is present	Short-term dataset is available. Limited documentation is present	All data and metadata are documented. Data provenance is documented and maintained	Sampling and analysis are manual and error-prone. No quality assurance or control is in place	Reasonable effort is made to ensure data quality. Limited product specific quality assurance is in place	Product information is available in metadata	Dataset integrity is not assured. No backup or recovery plan is in place



Stewardship Maturity Rating for NOAA's Product A – A Notational Example

Dataset Information: URL: Goes Here  
Dataset POC: Name & e-mail Here

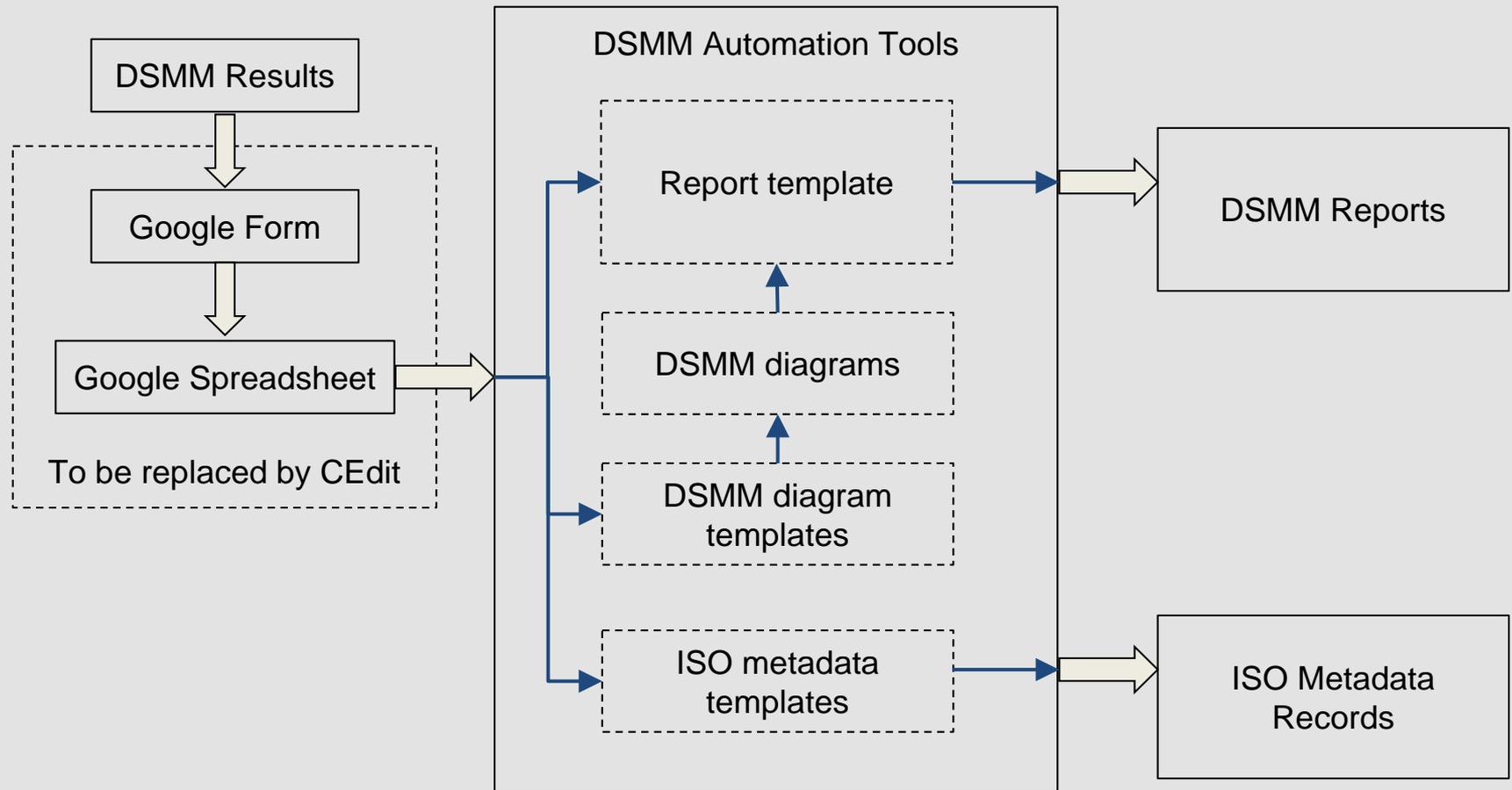
Category	Level 1	Level 2	Level 3	Level 4	Level 5
Preservability	★	★	★	★	★
Accessibility	★	★	★	★	★
Usability	★	★	★	★	★
Production Sustainability	★	★	★	★	★
Data Quality Assurance	★	★	★	★	★
Data Quality Control/Monitoring	★	★	★	★	★
Data Quality Assessment	★	★	★	★	★
Transparency / Traceability	★	★	★	★	★
Data Integrity	★	★	★	★	★

NOAA Technical Report NESDIS XXX  
doi: 10.7289/XXXXXXX

### Data Stewardship Maturity Report for Dataset Title



# Automation of DSMM report generation



# ISO Metadata Example (About DSMM)

## Conceptual Consistency

**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](https://doi.org/10.6084/m9.figshare.1211954)

Template  
Content



# ISO Metadata Example (Results)

## Conceptual Consistency

**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**

Results

### Conformance Result

**Explanation:** Data Stewardship Maturity Assessment was (etc...)

**Pass:** (inapplicable)

**Reference:** NOAA National Centers of Environmental Information. Stewardship Maturity Matrix (SMM) for NOAA Climate Data Record (CDR) of Monthly Outgoing Longwave Radiation (OLR), Version 2.2-1. 2016. DOI.

Results

# Example Metadata Snippet

```
</gmd:citedResponsibleParty>
</gmd:CI_Citation>
</gmd:evaluationProcedure>
<gmd:dateTime>
  <gco:DateTime>2016-06-27T00:00:00</gco:DateTime>
</gmd:dateTime>
<gmd:result>
  <gmd:DQ_QuantitativeResult>
    <gmd:valueUnit gco:nilReason="inapplicable" />
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#DataQualityAssessment" codeListValue="intermediate">intermediate</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#Accessibility" codeListValue="advanced">advanced</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#DataQualityControlMonitoring" codeListValue="minimal">minimal</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#ProductionSustainability" codeListValue="advanced">advanced</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#DataIntegrity" codeListValue="intermediate">intermediate</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#Preservability" codeListValue="advanced">advanced</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#TransparencyTraceability" codeListValue="intermediate">intermediate</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#Usability" codeListValue="advanced">advanced</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
    <gmd:value>
      <gco:Record>
        <gmd:CodeListValue codeList="http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/MM_StewCodelists.xml#DataQualityAssurance" codeListValue="advanced">advanced</gmd:CodeListValue>
      </gco:Record>
    </gmd:value>
  </gmd:DQ_QuantitativeResult>
</gmd:result>
</gmd:DQ_ConceptualConsistency>
</gmd:report>
</gmd:DQ_DataQuality>
```



# Recent Activities

- *OneStop* Team completed more than 660 assessments and associated documentation/metadata to date.
- Documentation meeting the NOAA Institutional Repository requirements
- Embedded Data Quality section of collection level ISO metadata records with DSMM results



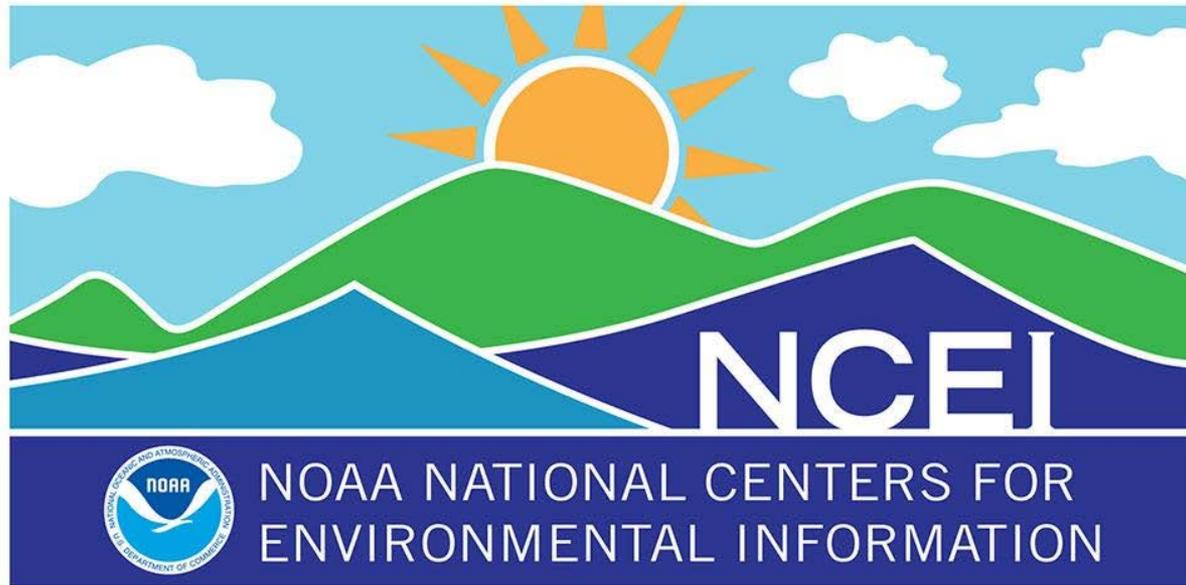
# Lessons Learned

- Engaging subject matter experts is imperative for completing and validating the DSMM results
- DSMM assessments take time and patience
- Automate as much as possible!



# Acknowledgements

- Members of the *OneStop* team and Product Stewards either participated in the implementation of the DSMM or provided information on practices applied to the datasets.
- This implementation is sponsored by both the NOAA OneStop Program and NCEI Data Stewardship Division.
- G. Peng is supported by NOAA under Cooperative Agreement NA14NES432003.



[www.ncei.noaa.gov](http://www.ncei.noaa.gov)  
[www.climate.gov](http://www.climate.gov)



NCEI Climate Facebook: <http://www.facebook.com/NOAANCElclimate>

NCEI Ocean & Geophysics Facebook: <http://www.facebook.com/NOAANCEloceangeo>

NCEI Climate Twitter (@NOAANCElclimate): <http://www.twitter.com/NOAANCElclimate>

NCEI Ocean & Geophysics Twitter (@NOAANCElocngeo): <http://www.twitter.com/NOAANCElocngeo>

