Assessing Climate Data Record Transparency and Maturity
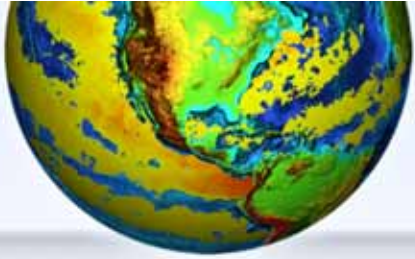
John J. Bates

Chief, Remote Sensing Applications Division

National Climatic Data Center, NOAA/NESDIS
April, 2011

Overview

- Climate variability and change are having profound effects on society

- Society deserves full and open access to the data and methods used to produce climate products

- Scientists often use nomenclature and methods that are difficult for non-specialists, and sometimes even specialists, to understand

- Scientists must do a better job in standardizing their nomenclature and methods, and more clearly communicate with the public
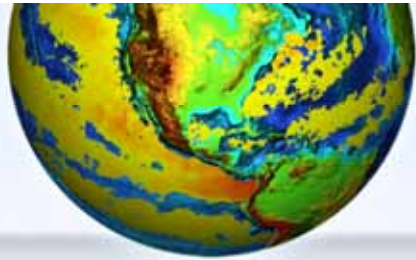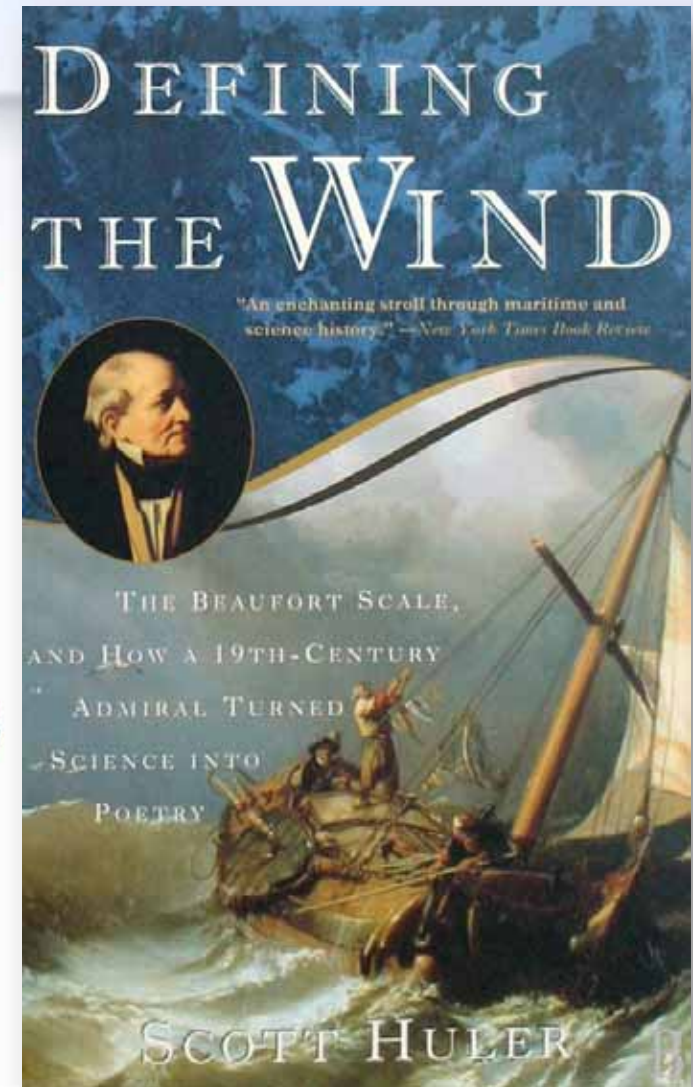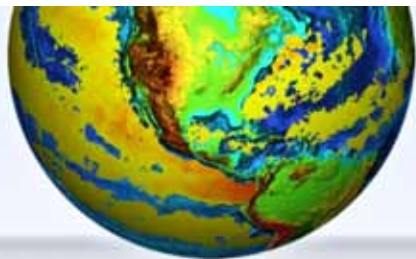
Motivation

What is at stake?

- History shows that weather observations did not become useful for society until a lexicon was agreed to

  - ✓ The Beaufort scale did this for wind climatology and maritime commerce in the 19th century

- For The Climate Service to benefit society, it must adopt a lexicon that sets expectations for openness, process and transparency that are accessible to the public

  - ✓ How might we define a climate record lexicon useful to both scientists and the general public in the 21st century?


DEFINING THE WIND

"An enchanting stroll through maritime and science history." —*New York Times Book Review*

THE BEAUFORT SCALE, AND HOW A 19TH-CENTURY ADMIRAL TURNED SCIENCE INTO POETRY

SCOTT HULER

Business Practices

Steps to long-term monitoring

- Over the last 20-30 years many investigators have developed methods for seaming together observations with evolving coverage and accuracies

- From these experiences, common elements are emerging on how climate scientists do business
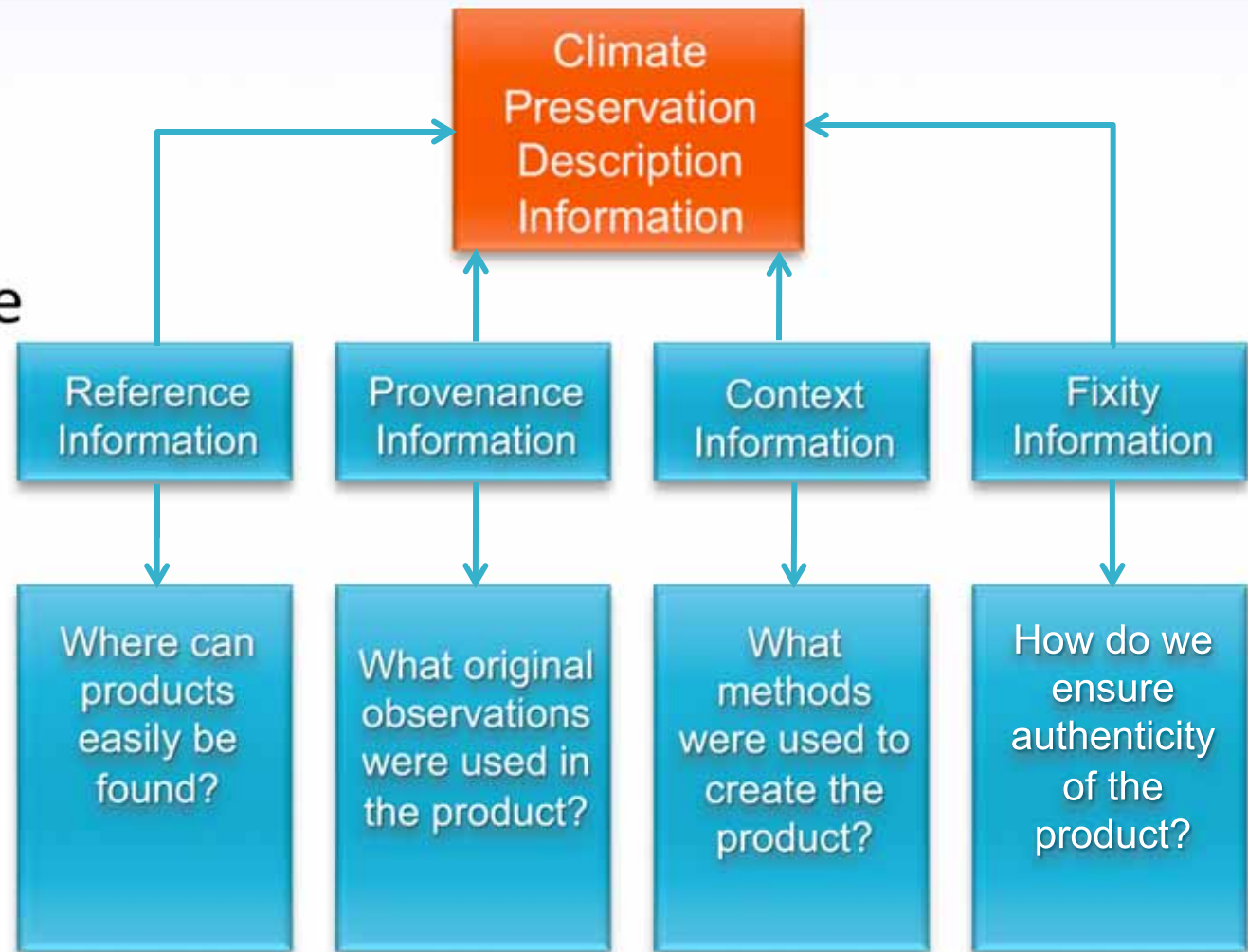
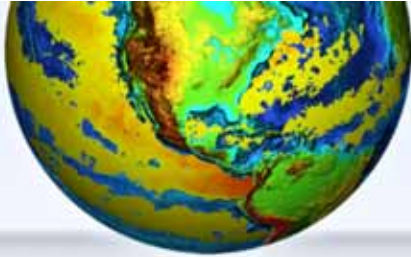## How do we capture and make available these business practices?

60°S-60°N Average SST
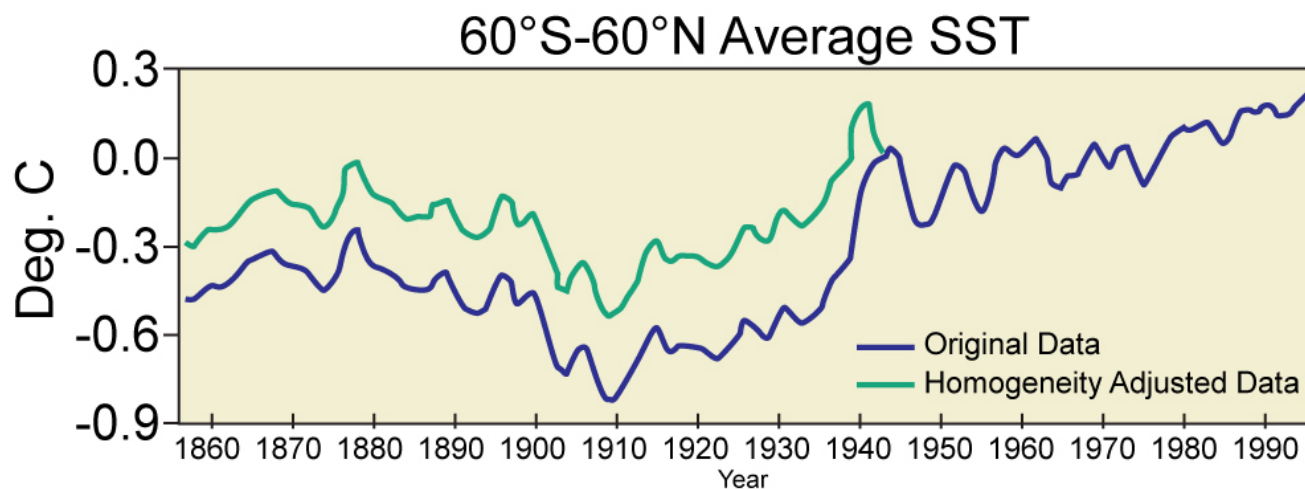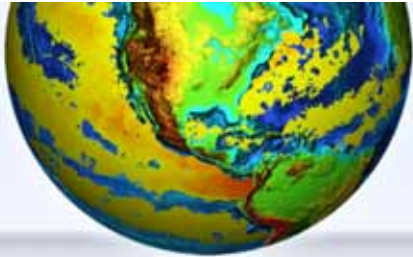
Common Climate Observations
Business Practices

What common preservation information do we need and how do we capture and communicate this?

Climate Preservation Description Information

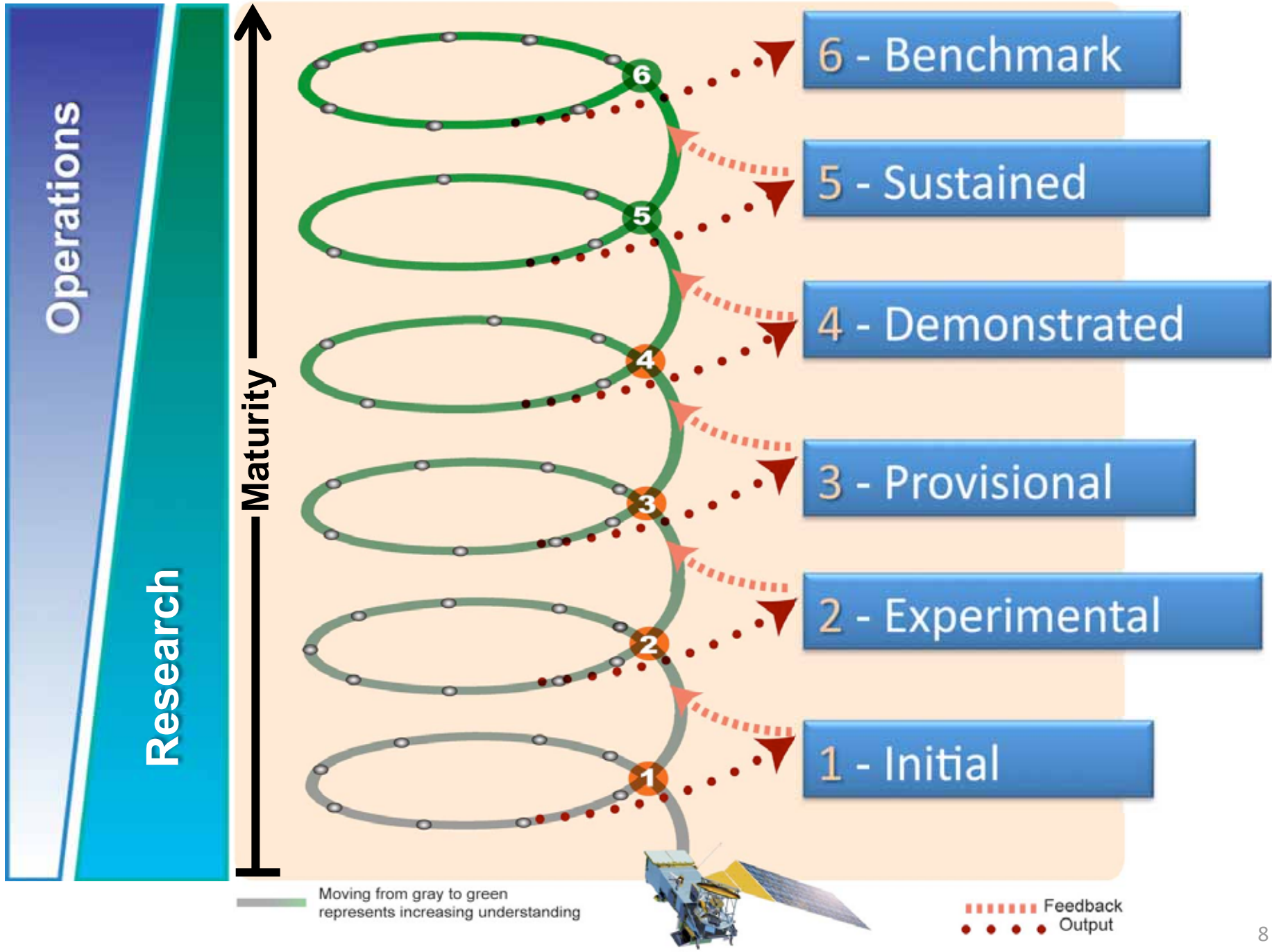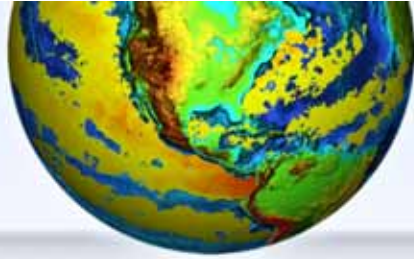| Reference Information	Provenance Information	Context Information	Fixity Information
Where can products easily be found?	What original observations were used in the product?	What methods were used to create the product?	How do we ensure authenticity of the product?

Do We Need Observation Ingredients and Nutritional Labels?

Preservation Description Information

- Reference Information → Store Aisle Shelf
- Provenance Information → Ingredients
- Context Information → Nutrition Facts
- Fixity Information → Tamper Proof Seal

Maturity Matrix

Where can products easily be found?

What original observations were used in the product?

What methods were used to create the product?

How do we ensure authenticity of the product?

Climate Portal www.climate.gov

Digital Signature

Let's define a Maturity Matrix (1=low; 6=high) that sets expectations and assesses progress

Collection Methods	Algorithm stability	Metadata & QA	Documentation	Validation	Public Release	Science & Applications
How was the data collected, sensors, surveys, etc.?	Are algorithms under configuration management and how mature?	How full and complete are the metadata and quality assessment?	Is the Operational Algorithm Description full, complete, and peer reviewed?	How complete is the validation?	Are the data, algorithms and software open and available to the Public?	How extensive is the peer reviewed literature and how varied are the applications?

CDR Evolution

Operations

Research

Maturity

6 - Benchmark

5 - Sustained

4 - Demonstrated

3 - Provisional

2 - Experimental

1 - Initial

Moving from gray to green represents increasing understanding

Feedback
Output

How Do We Prioritize Climate Record Work?

- We can group by cycles –
  Water Cycle, Carbon Cycle, etc.

- We can group by forcing and feedbacks

- We can group by Societal Benefit Areas

- Regardless of the approach, we need some way to assess and communicate easily our progress towards maturity
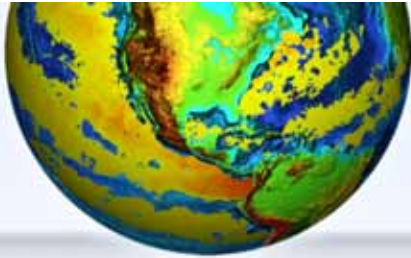
Using the Maturity Matrix to Assess Progress

- NOAA's Climate Data Record (CDR) Program is working with scientists on the routine production of climate information
- A self assessment by those scientists provides a first measure of how the climate community is doing in meeting criteria for openness, process and transparency
- Results show moderate levels of maturity and more work needed in particular on metadata and documentation

Self-assessment of a single CDR
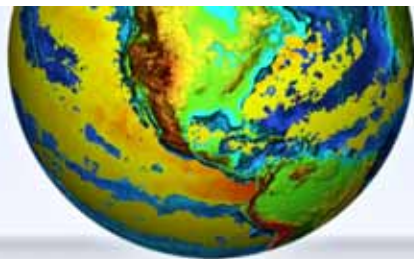

CDR Program Maturity Matrix Statistics
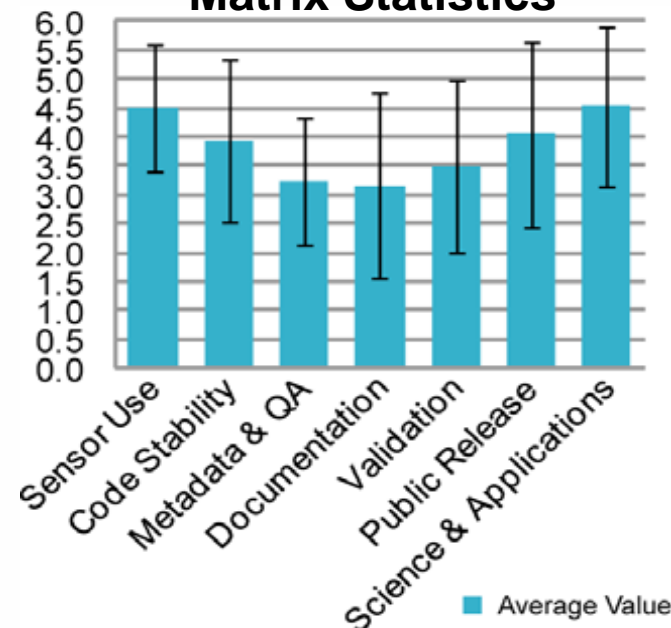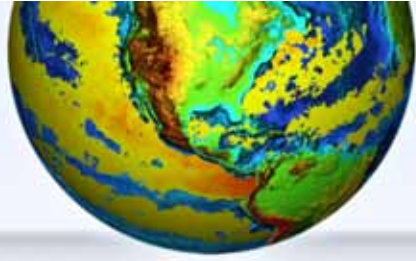
What is the Role of IT in Enabling Transparency and More Widespread Use of Climate Data?

- The Maturity Matrix sets expectations, but needs enabling information technology to be broadly useful

- Climate Services, to be relevant to society, needs to be accessible to both expert and non-expert communities

- We need what has been dubbed '*Knowledge provenance*'* - that is, for the content of the Maturity Matrix to be truly useful it must be enriched with semantics and semantically-aware tools

* The term 'knowledge provenance' was coined by Peter Fox Rensselaer Polytechnic Institute

Semantics – Enabling Data Transparency

Slide Courtesy Peter Fox RPI

12
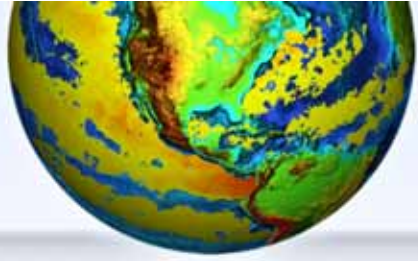
Conclusions

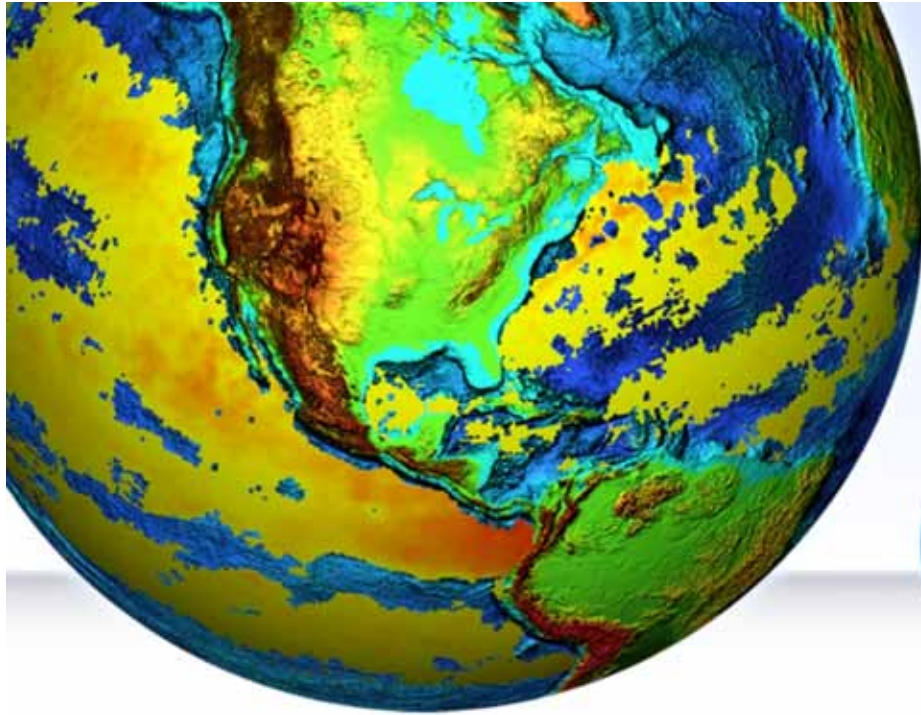- Climate Scientists must adopt a common vocabulary and lexicon

- The Maturity Matrix provides a basis for information preservation, expectations, and a metric for progress to completeness

- The Maturity Matrix concept needs to be enabled through the use of semantics to be useful to a broad range of interdisciplinary scientists and policy makers

Thank You...

Questions?