

Assessing Climate Data Record Transparency and Maturity

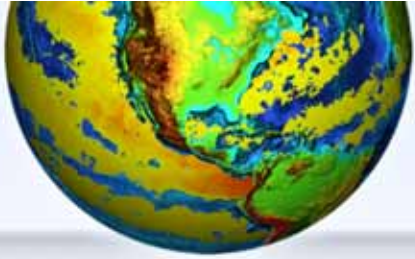


John J. Bates

Chief, Remote Sensing Applications Division

National Climatic Data Center, NOAA/NESDIS

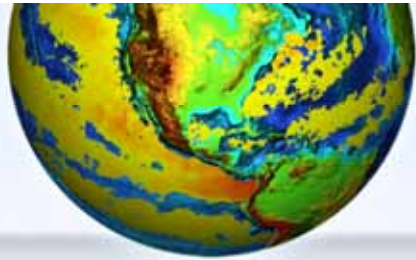
April, 2011



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

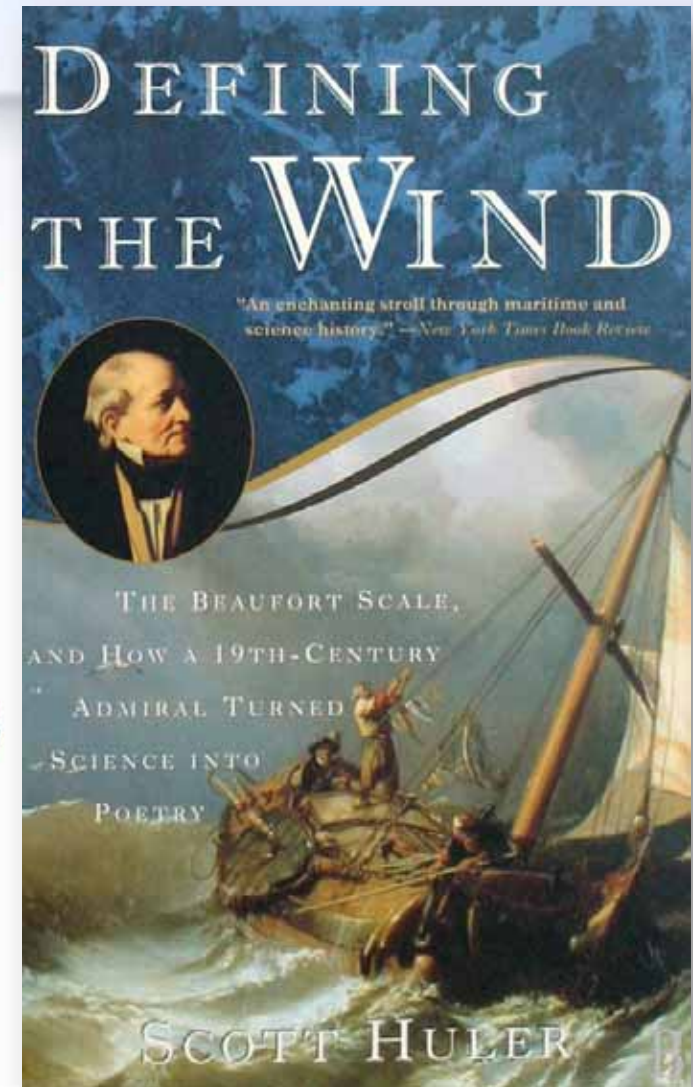


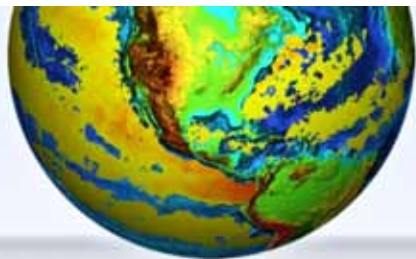


Motivation

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?



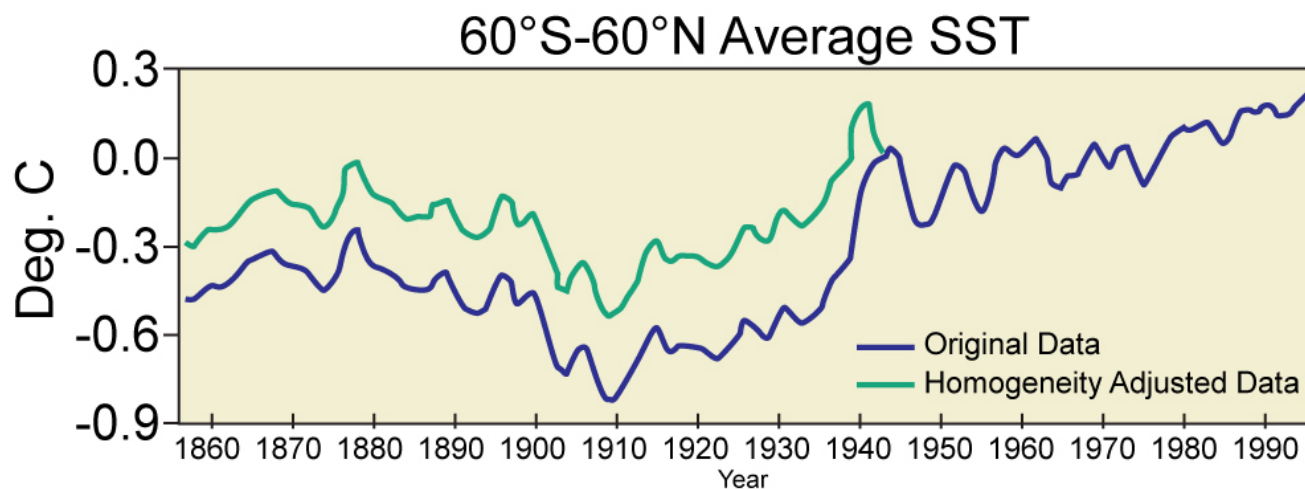


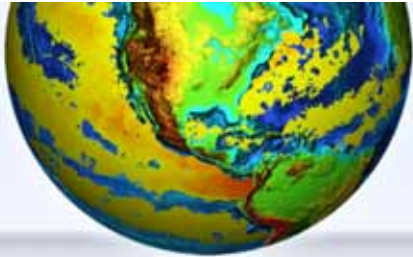
Common Climate Observations Business Practices

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

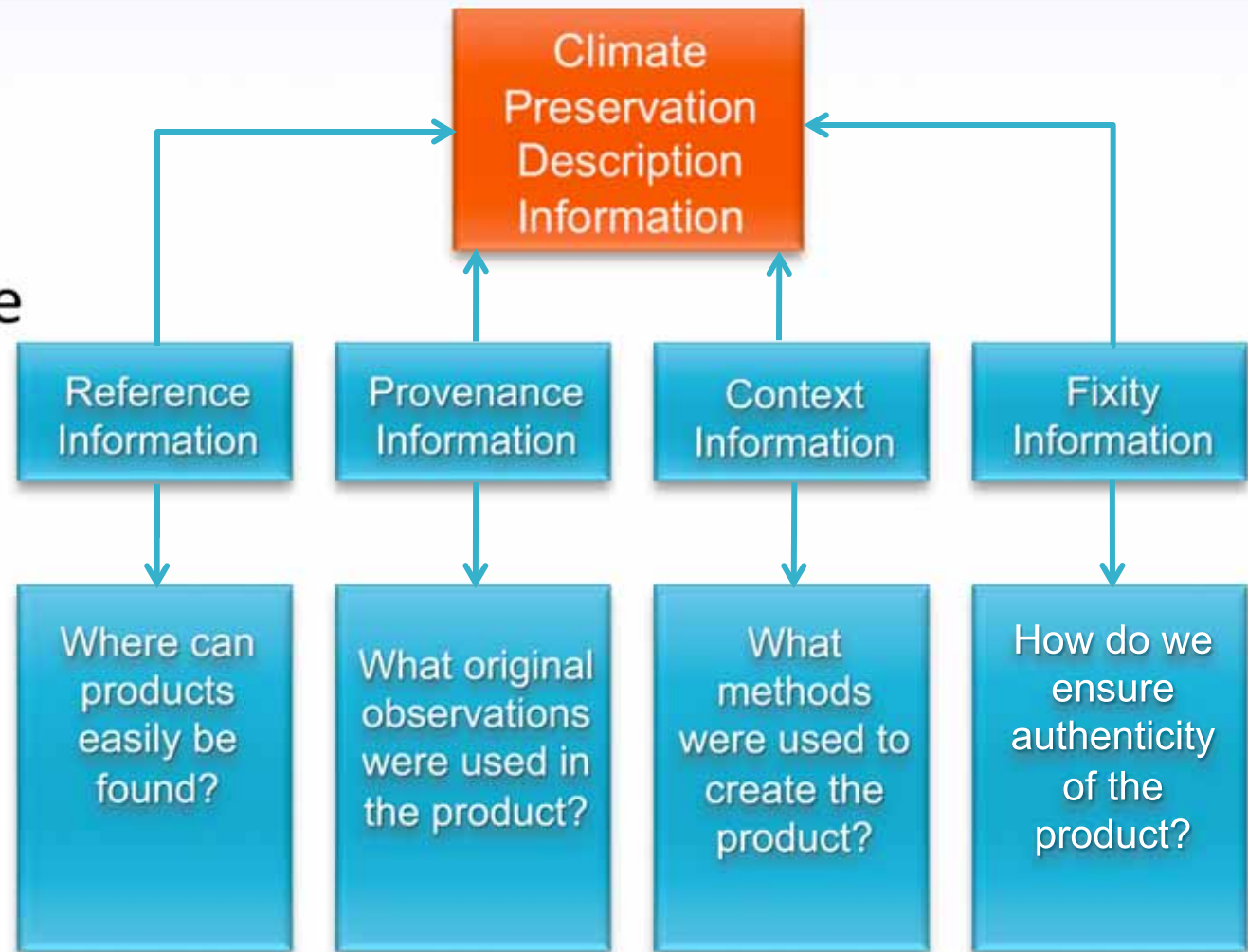
How do we capture and make available these business practices?

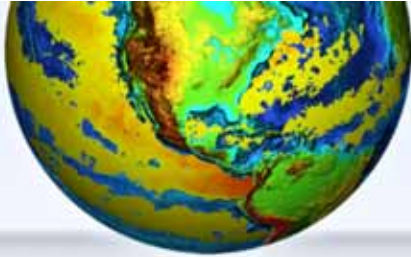




Common Climate Observations Business Practices

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





Do We Need Observation Ingredients and Nutritional Labels?

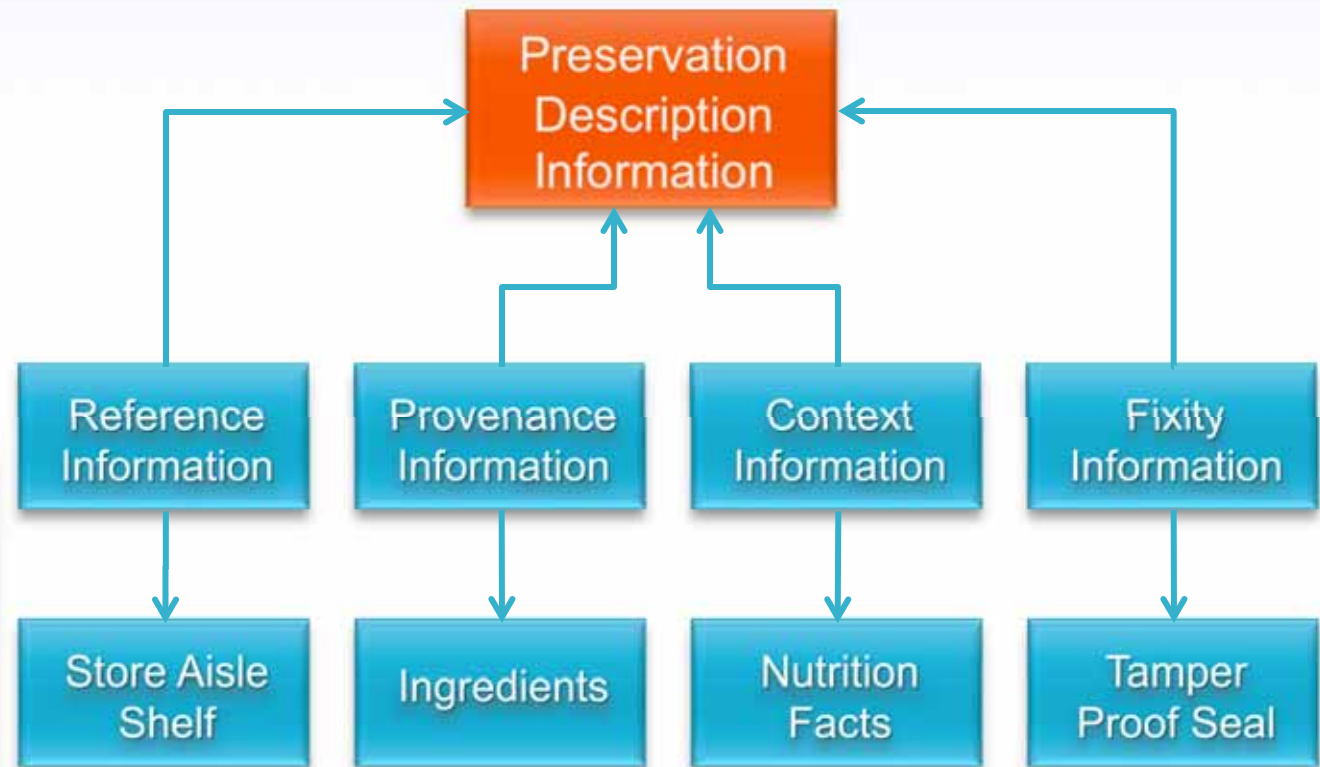


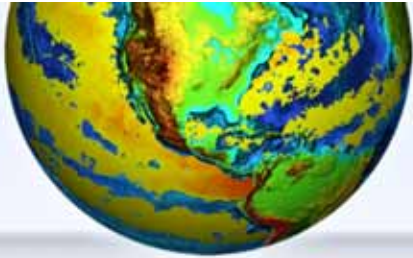
Ingredients: Tapioca Syrup*, Dark Chocolate Flavored Coating (sugar*, palm kernel oil, cocoa powder*, cocoa butter*, soy lecithin*, nonfat milk*, vanilla*), Peanuts*, Granola* (rolled oats*, brown rice syrup*, sugar*, sunflower oil*, sea salt, baking soda, vanilla*), Crisp Rice* (rice flour*, sugar*, sea salt, malt extract*, annatto color), Cashews*, Almonds*, Rice Maltodextrin*, Sea Salt, Sunflower Oil*, Soy Lecithin*, Natural Flavor. *Organic CONTAINS PEANUT, CASHEW, ALMOND, SOY, MILK; MAY CONTAIN WALNUT, PISTACHIO, PINE NUT, PECAN, MACADAMIA, HAZELNUT AND BRAZIL NUT INGREDIENTS.

Nutrition Facts

Serving Size	(35g)
Amount Per Serving	
Calories	160
Calories from Fat	70
% Daily Value*	
Total Fat 8g	13%
Saturated Fat 3g	14%
Trans Fat 0g	
Cholesterol 0mg	0%
Sodium 130mg	5%
Total Carbohydrate 20g	7%
Dietary Fiber 2g	7%
Sugars 10g	
Protein 3g	
Iron	2%
Not a significant source of vitamin A, vitamin C and calcium.	
*Percent Daily Values are based on a 2,000 calorie diet.	

CERTIFIED ORGANIC
 CERTIFIED ORGANIC BY THE WASHINGTON STATE DEPARTMENT OF AGRICULTURE IN ACCORDANCE WITH THE ORGANIC STANDARDS OF THE U.S. DEPARTMENT OF AGRICULTURE.





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?

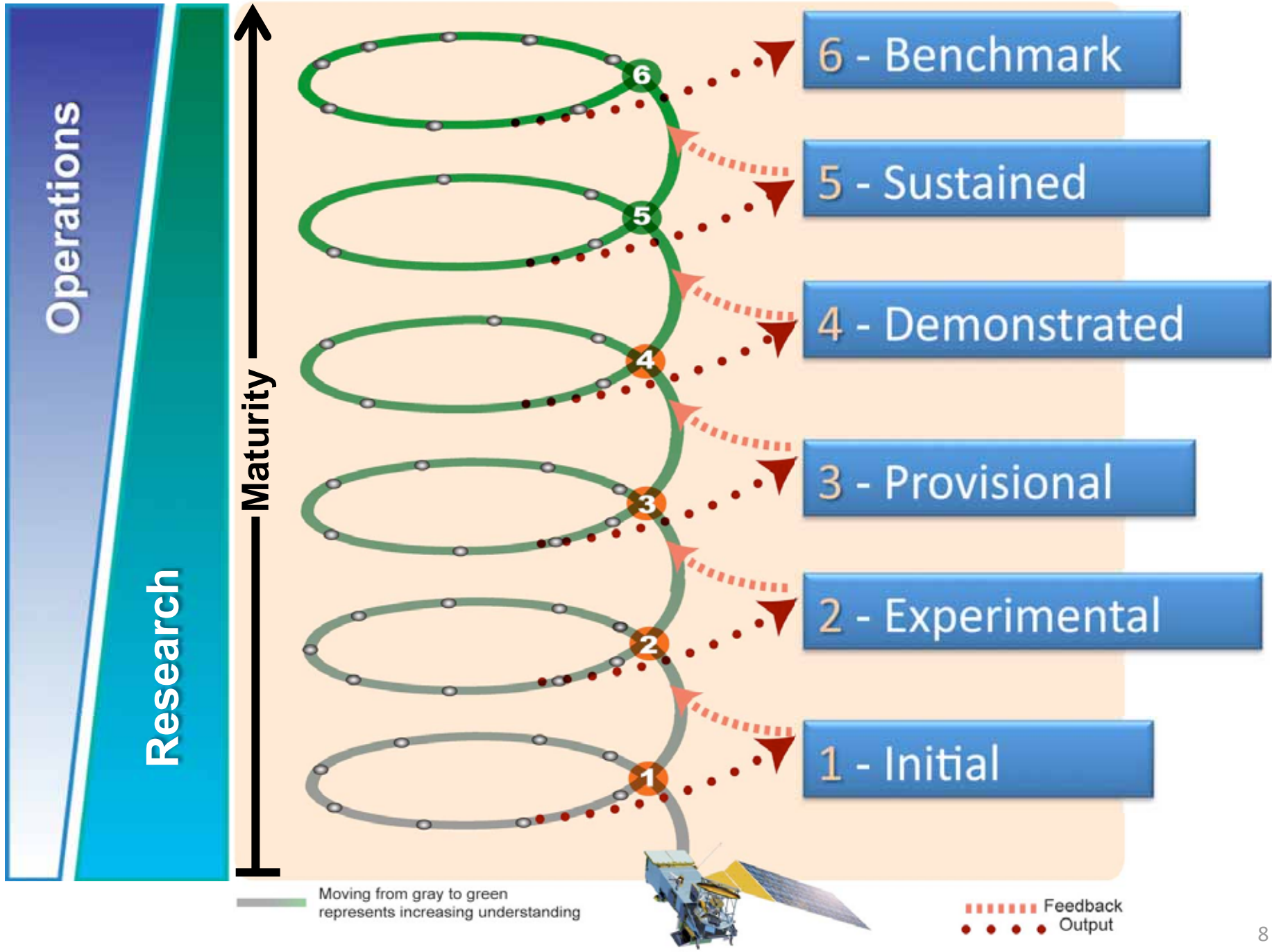
Climate Portal
www.climate.gov

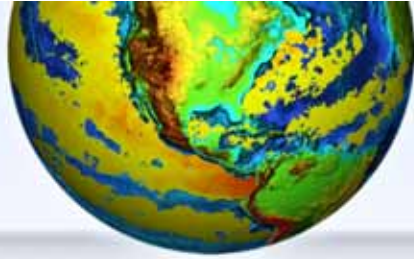
Digital Signature

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

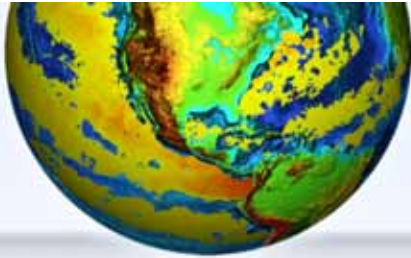




How Do We Prioritize Climate Record Work?

- 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





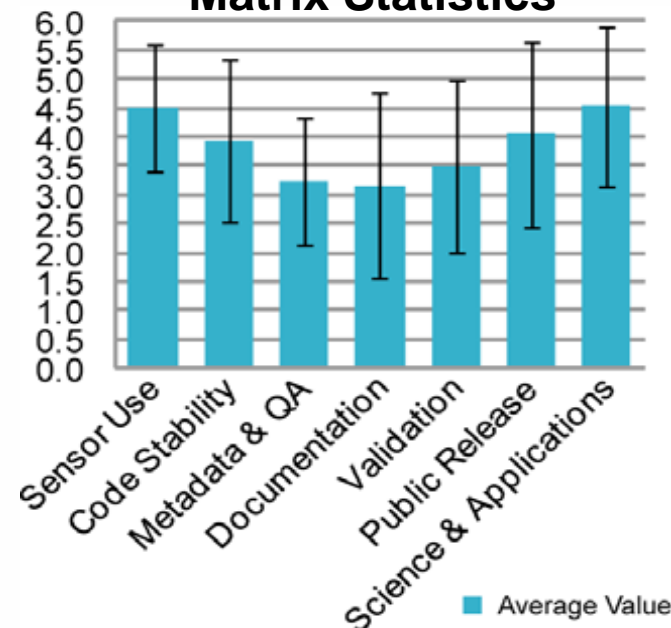
Using the Maturity Matrix to Assess Progress

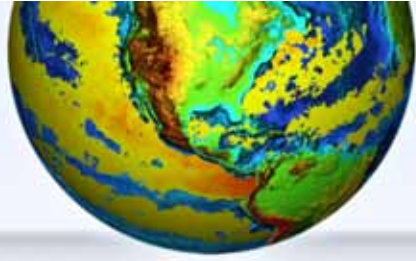
- NOAA's Climate Data Record (CDR) Program is working with scientists on the routine production of climate information
- A self assessment by those scientists provides a first measure of how the climate community is doing in meeting criteria for openness, process and transparency
- Results show moderate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

Maturity	Sensor Use	Algorithm stability	Metadata & QA	Documentation	Validation	Public Release	Science & Applications
1	Research Mission with limited period of record	Significant changes likely	Incomplete	Draft Operational Algorithm Description (OAD)	Minimal	Limited data availability to develop flexibility	Little or none
2	Research Mission with limited period of record	Some changes expected	Research grade (constraints)	OAD Version 1+	Uncertainty estimated for select locations/seasons	Some available but of unknown accuracy, certain required for use	Limited or ongoing
3	Research Mission with sufficient period of record	Minimal changes expected	Research grade (constraints), Meets international standards	Peer-reviewed OAD and product description	Uncertainty estimated over widely distributed time/locations by multiple investigators, differences understood	Data available but of unknown accuracy, certain required for use	Provisionally used in applications and assessments demonstrating positive value
4	Operational Mission with sufficient period of record	Minimal changes expected	Stable, Address performance tracking and reproducibility, Meets international standards	Public Operational Algorithm Description (OAD), Peer-reviewed product description	Uncertainty estimated over widely distributed time/locations by multiple investigators, differences understood	Data archived and available but of unknown accuracy, certain required for use	Operationally used in applications and assessments demonstrating positive value
5	All relevant research and operational missions, unified and coherent record documented across different sensors	Stable and reproducible	Stable, Address performance tracking and reproducibility, Meets international standards	Public OAD and Validation Plan, Peer-reviewed product and validation articles	Consistent uncertainties estimated over most environmental conditions by multiple investigators	Multi-mission record is archived and publicly available with associated uncertainty estimate	Used in published applications and assessments by different investigators
6	All relevant research and operational missions, unified and coherent record over complete series, record is consistent in monthly resolution following common schema	Stable and reproducible, homogeneous and published over budget	Stable, Address performance tracking and reproducibility, Meets international standards	Product, algorithm, validation, processing and metadata described in peer-reviewed literature	Observation strategy designed to reveal systematic errors through independent cross-checks, open inspection, and continuous interrogation	Multi-mission record is publicly available from long-term archive	Used in multiple published applications and assessments by different investigators
Comments for Maturity rating	FOR 1854 in process	Code documented throughout	FORC compliant	Algorithm techniques published in multiple papers	Comparisons made with equivalent products e.g. MURST and OST	Data archived and available on the server	Well published and referenced research
Avg rating = 3.9	Ship and buoy data consists of (CDR/CDR 4 in various sites after 2006)	Multiple papers published	User Manual available	Internal wiki page with overview, flow chart, white papers, and code descriptions	Comparisons with standards in e.g. GDR, CMC 1 but need others	Product updated monthly	Most recent version will need a paper after seeing the Office adjustment factors and new CDR4D release
	NOTE: pseudo calibration uses average statistical adjustment factors no instrument	Source code is packaged and deployable		Source code is packaged and deployable	Comparisons among previous versions		

Self-assessment of a single CDR

CDR Program Maturity Matrix Statistics

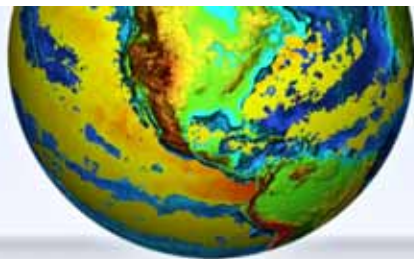




What is the Role of IT in Enabling Transparency and More Widespread Use of Climate Data?

- 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

People

Agency Policy Makers

System Scientists

Politicians

Decision-level semantic mediation: high-level vocabularies that facilitate policy-level decision-making

Integrated Applications

Inter-disciplinary Data Visualization Apps

Semantic interoperability

Integration Frameworks & Methodologies

Eco & other system Assessment Apps



Application-level semantic mediation: mid-level vocabularies that facilitate the interoperability of system models and data products

Software, Tools & Apps

Discipline-specific model(s)

Semantic interoperability

Dataproduct Generator

Information / Science Apps



Semantic query, hypothesis and inference

Query, access and use of data

Data-level Semantic mediation: lower-level vocabularies applied to each data source for a specific science domain of interest

Data Repositories

Federal Repository

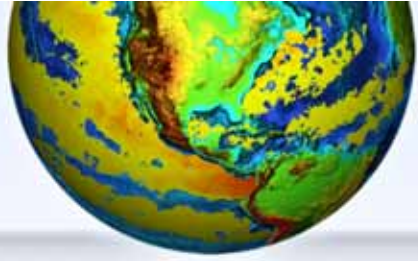
Commercial Database

Researcher Private Database

Other Data Sources

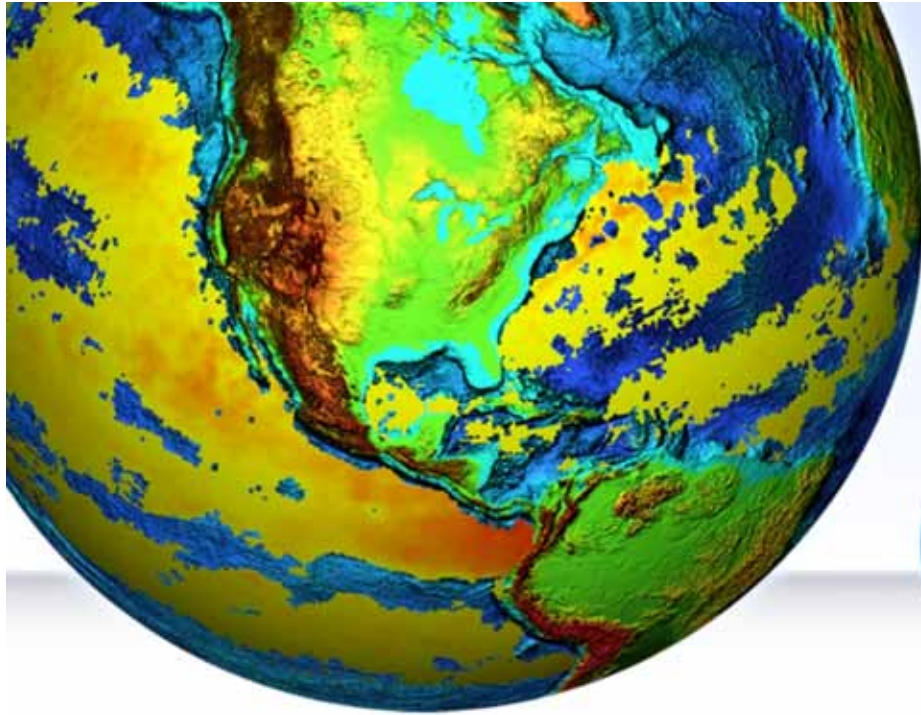
Metadata, schema, data
... ..

Slide Courtesy Peter Fox RPI



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



THE CLIMATE SERVICE

Thank You...

Questions?