Archive Management
The Missing Component

Howard J. Diamond, John J. Bates
David M. Clark, Robert L. Mairs

NOAA/NESDIS

IEEE/NASA Mass Storage Conference

Presented by:
John A. Jensen, NOAA/NCDC
April 7-10, 2003; San Diego, CA

Archive Management
What's Unique in NOAA?

1. Access to ALL Data & Information by ALL
2. Data & Information is held in "Perpetuity"
3. Soonest Possible Access to New and Historical Data & Information.
4. Specific Data & Information
5. Federal Records Center – NARA Standards

Today's Customers - Users

- "Want information and answers to specific questions rather than simply access data."
- "No longer content to wait days for data and information."
- "Demand on-line inventories, search, browse, ordering, and immediate electronic transfer."
- "New User Groups supporting wide range of decision making and rapid responses to immediate needs."

Three Components of Archive Management

"**Data & Information Stewardship**"

1. Customer Services
2. Scientific Stewardship
3. Information Technology (IT) Infrastructure

Customer Services

- Data Quality and Continuity
- Ingest
  - Real Time (minutes to hours)
  - Near Real Time (hours to a day)
  - Delayed (days, weeks, months)
- Access
  - On-Line: Disk (WWW/Internet, NGI, etc.)
  - Near On-Line (robotics)
  - Off-Line (paper, microfilm)

Customer Services

On Demand Specific Data & Information:

New, Recent, and Historical

- On-Line Comprehensive Inventory, Search, Order, Delivery
- On-Line - Now: Disk
- Near On-Line: Robotics
  - Immediate Retrieval and Staging
  - Soonest Retrieval and Staging
- Off-Line: Paper & Microfilm records

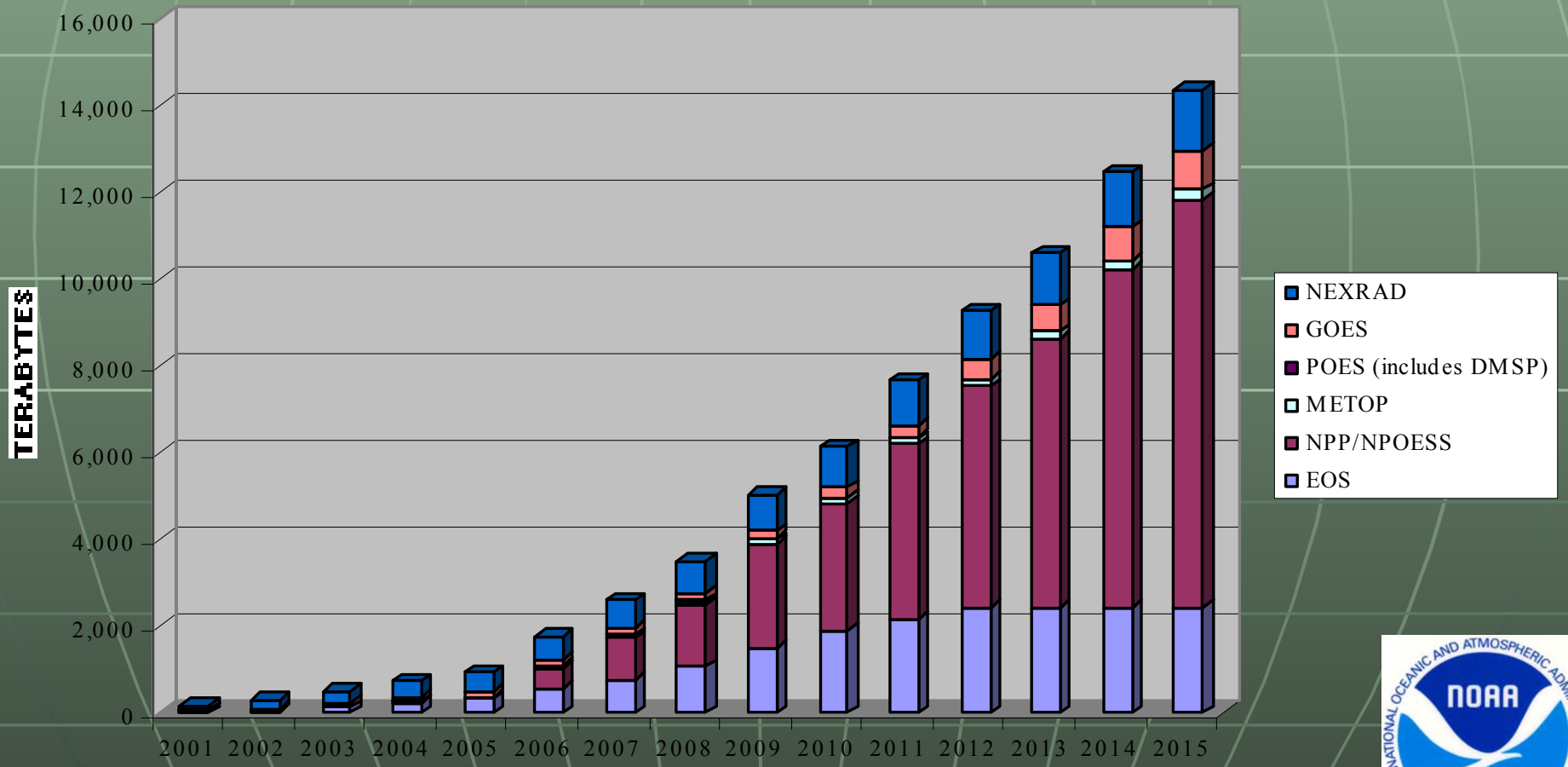
Customer Services

- Data Mining: New, Recent, and Historical
  - During Ingest Phase
  - Recent Data
  - Historical (archived) Data
- Data Fusion: New, Recent, and Historical
  - During Ingest Phase
  - Recent Data
  - Historical (archived) Data
- Off-Line: Paper & Microfilm

"Data & Information Stewardship"

1. Customer Services
2. Scientific Stewardship
3. Information Technology (IT) Infrastructure

Cumulative Major Systems Archive Growth (not including backup)



Scientific Stewardship

Why Scientific Stewardship?

To meet the Challenge of Capitalizing on:

True Potential Value and Use
of Information and Knowledge.

Scientific Stewardship

Characterized as:

"Maintaining the scientific integrity and long term utility of climate records"

through:

- **Monitoring**
- **Improving Quality**
- **Extraction of Select Key Parameters**

Scientific Stewardship

" **A Data Management Discipline**" **to ensure:**

- **Quality and Utility of data and information beyond initial and immediate use.**
- **Meaningful and Derived Information.**
- **Practical Application.**

Scientific Stewardship

"**A Data Management Process**" encompasses:

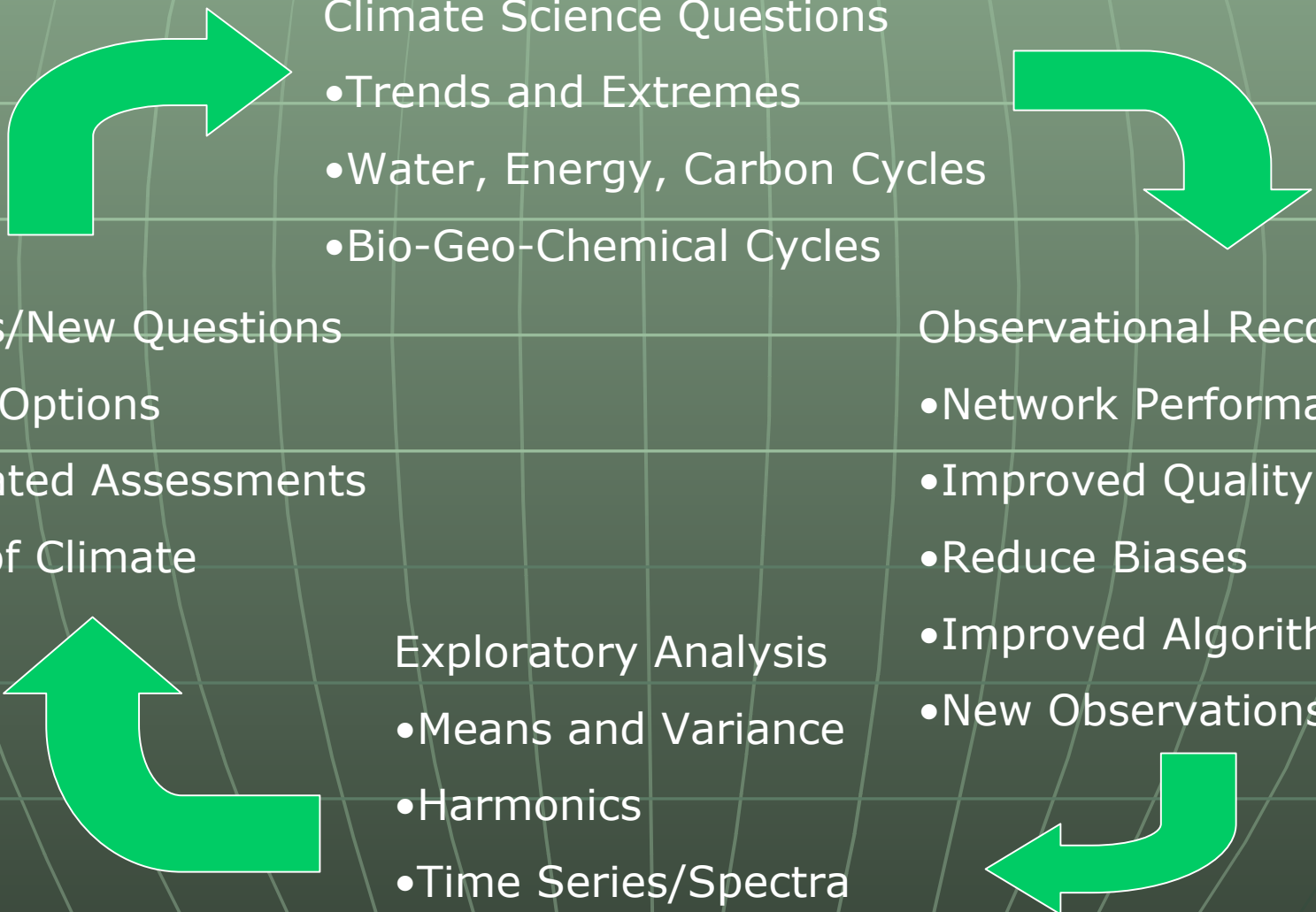
- Transformation of Data to Meaningful Information
- Information to Knowledge
- Knowledge to Understanding

To Enhance the formulation of

Sound Economic and Environmental Planning, Policies, and Decisions

Answer Pressing Science Questions

Closing the Loop on End-to-End Use of Global Observations



NOAA Scientific Stewardship

Five Priorities:

1. Ensure Observing System Quality.
2. Provide Common IT Infrastructure Support.
3. Develop "Climate" Processing System.
4. Document Earth System Variability.
5. Enable and Facilitate Future Research.

NOAA Scientific Stewardship

Five SDS Functions:

1. Provide real time automated monitoring.
2. Data quality & processing, i.e., *Data Mining* and *Data Fusion* (merging).
3. Produce authoritative records (CDRs).
4. Data processing and storage methods and procedures.
5. Data archaeology – data rescue.

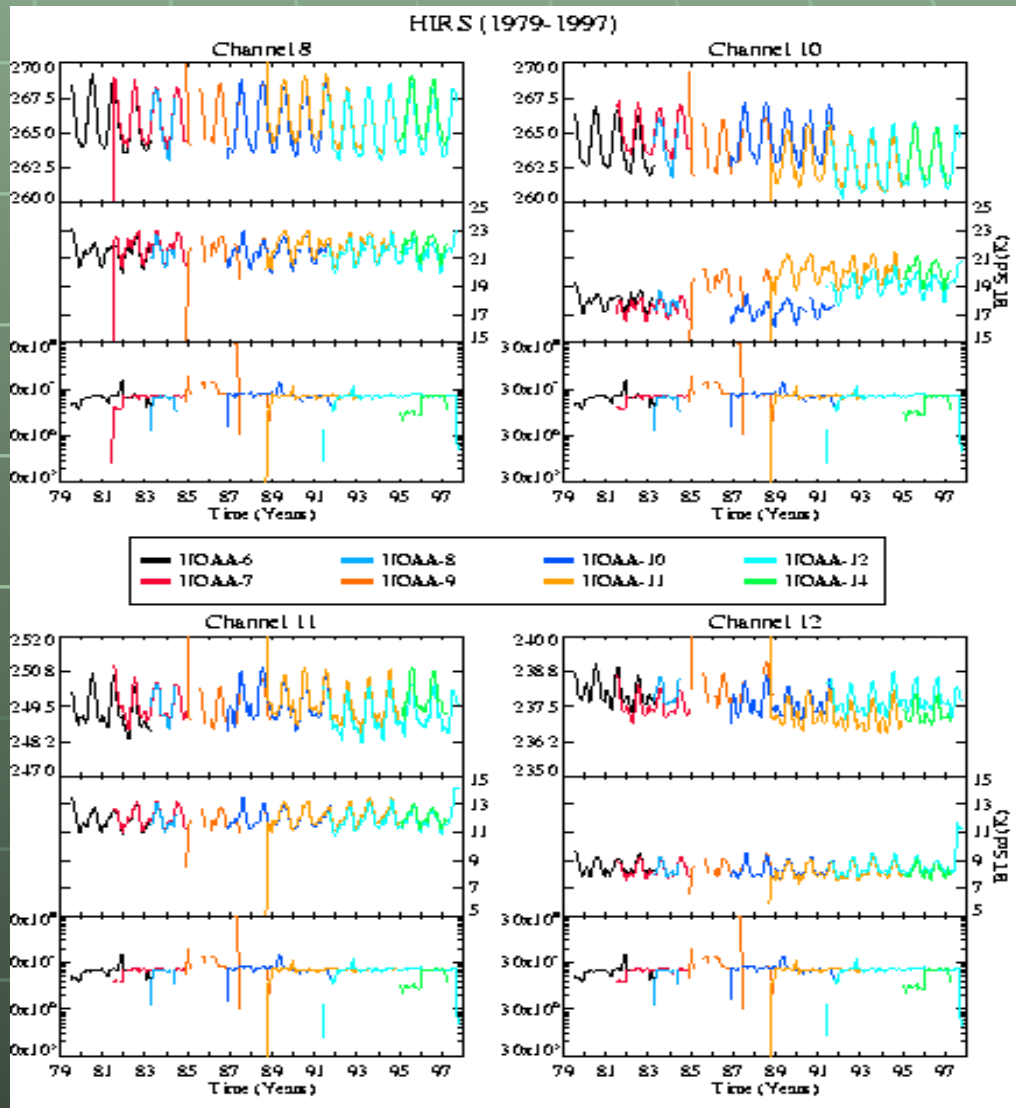
Scientific Stewardship

Monitor Observing System Performance

- Identify data quality problems early before they get big & in the archive.
- Take corrective action.
- Systematically improve observing system quality at ingest, more before archiving, & reprocessing.

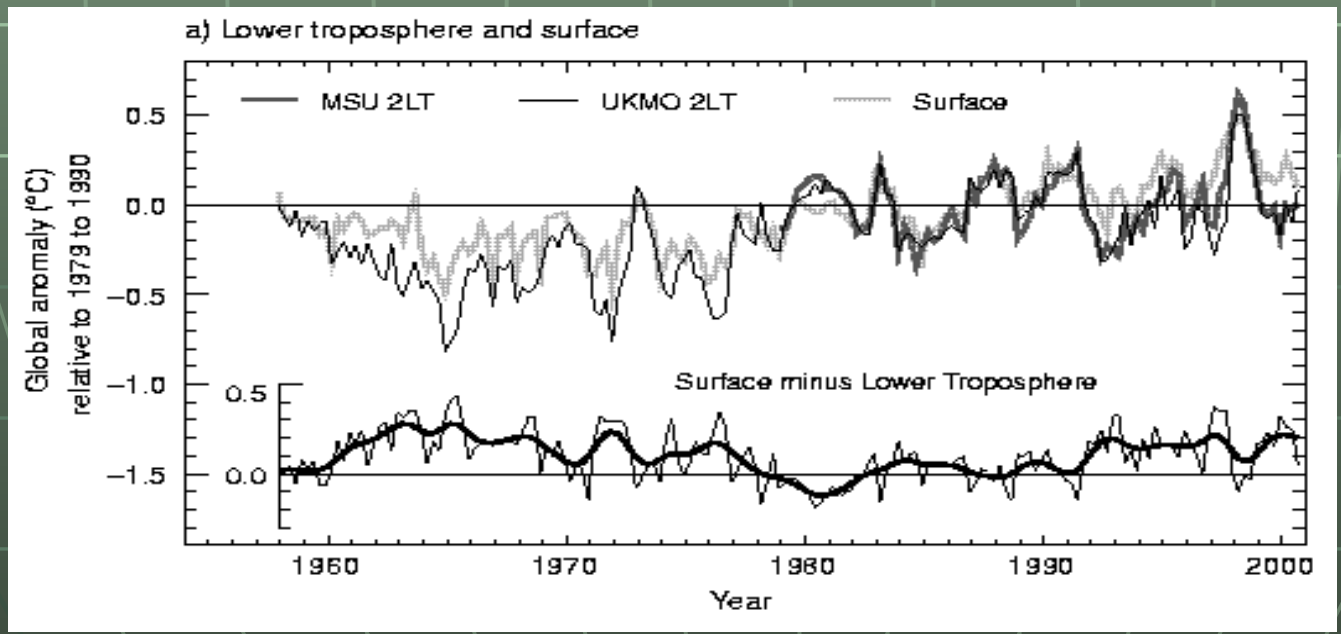
Monitoring the Observing System (Health of the Network)

- Continuously monitor channel vital signs.
- Correct systematic biases.
- Normalize variance by sensitivity.



End-To-End Example
MSU Lower Tropospheric Temperature

- Physical differences between Surface and MSU.
- Controversy in satellite-to-satellite "bias adjustments".
- Empirical versus physical bias adjustment.



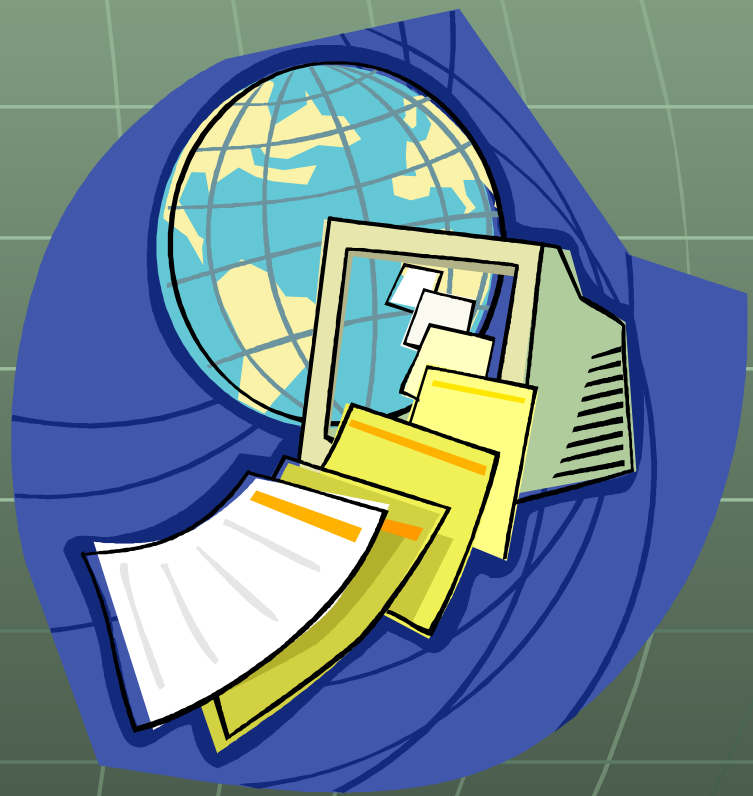
Scientific Stewardship

Provide IT Hardware & Software Support

- Assure flexible and efficient use of resources.
- Reduce duplication.
- Adapt quickly to new IT developments.

Comprehensive Large Array data Stewardship System (CLASS)

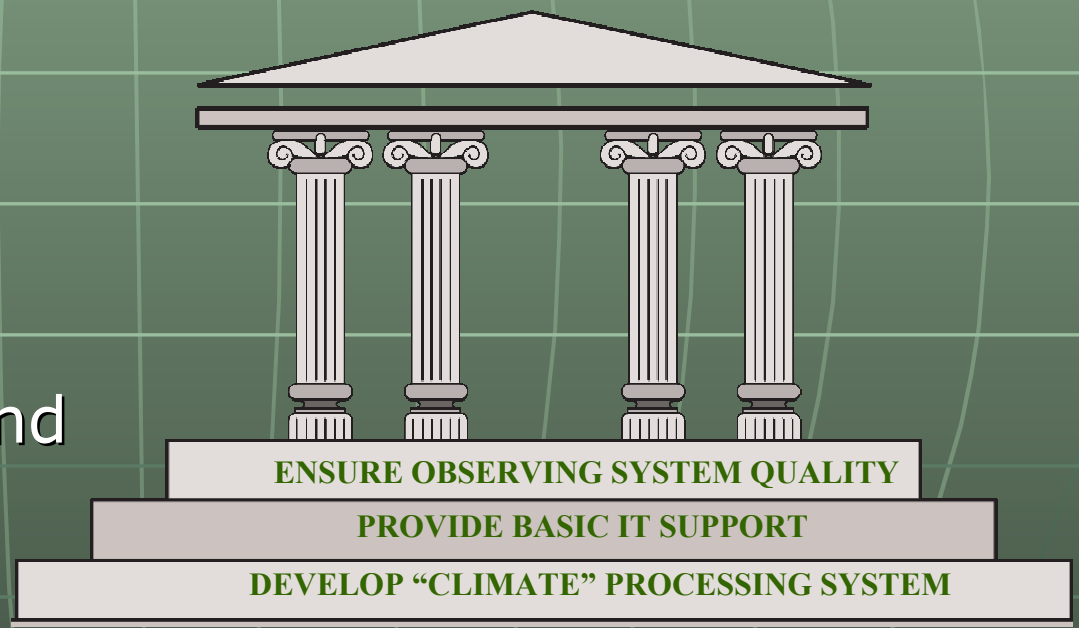
- A re-engineering effort.
- A phased implementation of major parts.
- Main portal for environmental data entrusted to NESDIS & NOAA stewardship.
- Ingest, Access, & Archive for large array data sets.



Scientific Stewardship

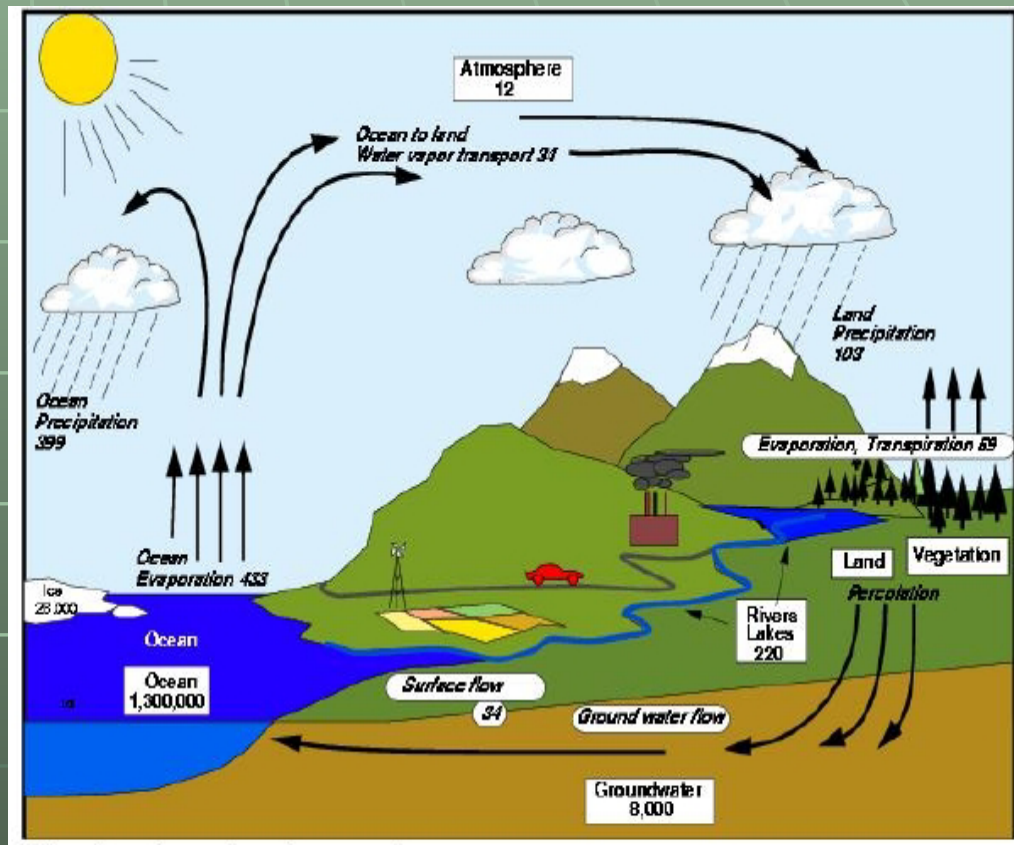
Develop "Climate" Processing System

- Remote (satellites) sensing and In-situ observations.
- Expert teams inside and outside NOAA.



Develop Climate Processing System

- NOAA-wide involvement.
- Cooperative Institute programs.
- NASA Cooperative Agreements.
- Grants and Contracts program with academia and industry.

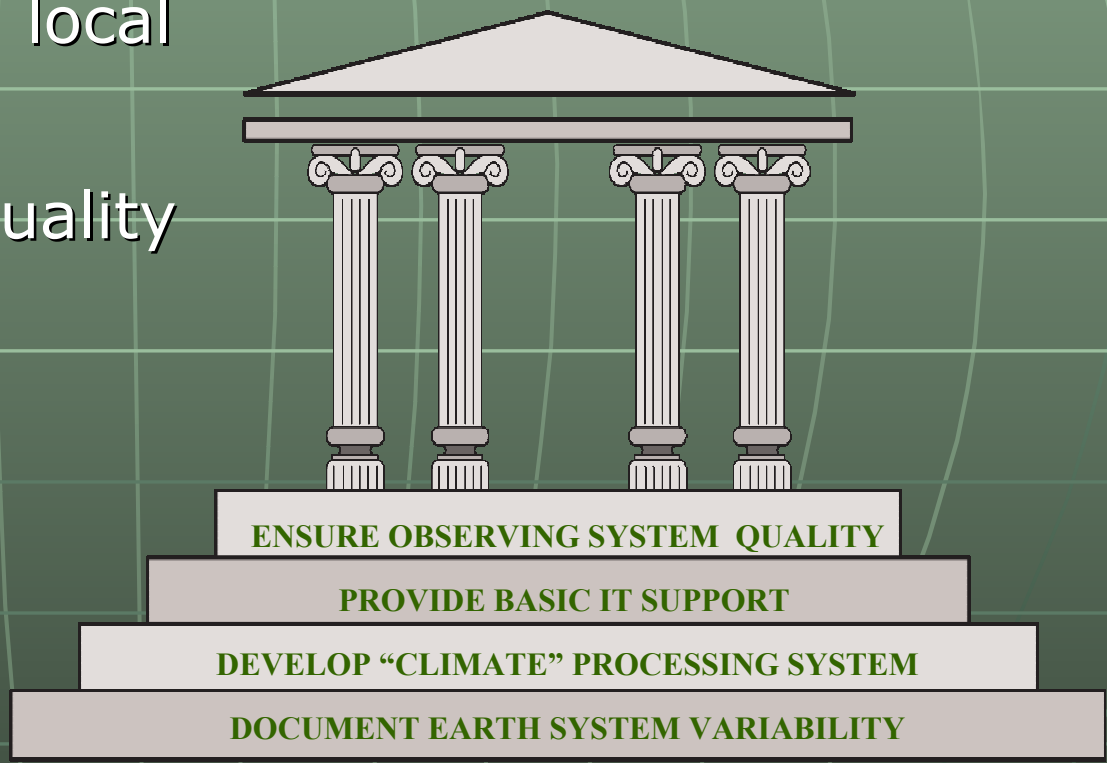


Hydrological cycle.
Units are thousand cubic km for storage and thousand cubic km/year for exchanges

Scientific Stewardship

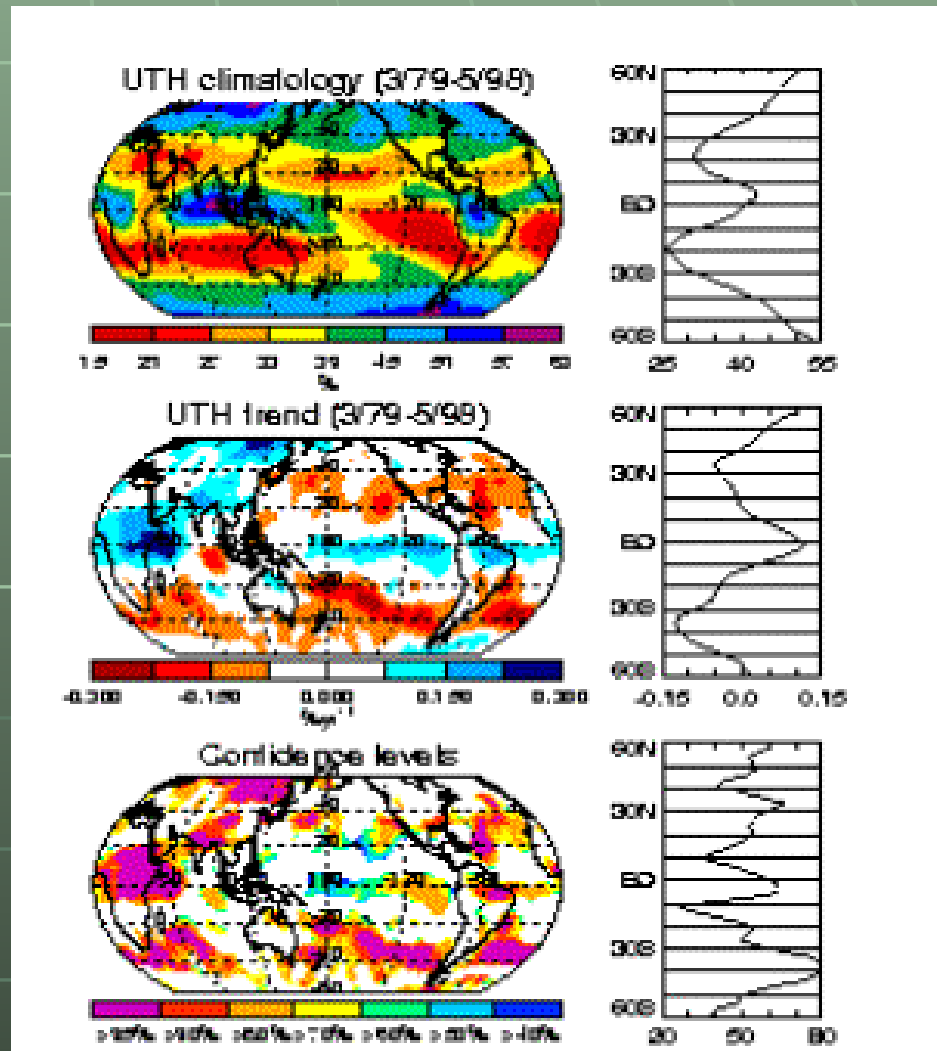
Document Earth System Variability

- Global, regional, and local scales.
- Build and maintain quality climate data bases.
- Build Climate Data Records (CDRs).



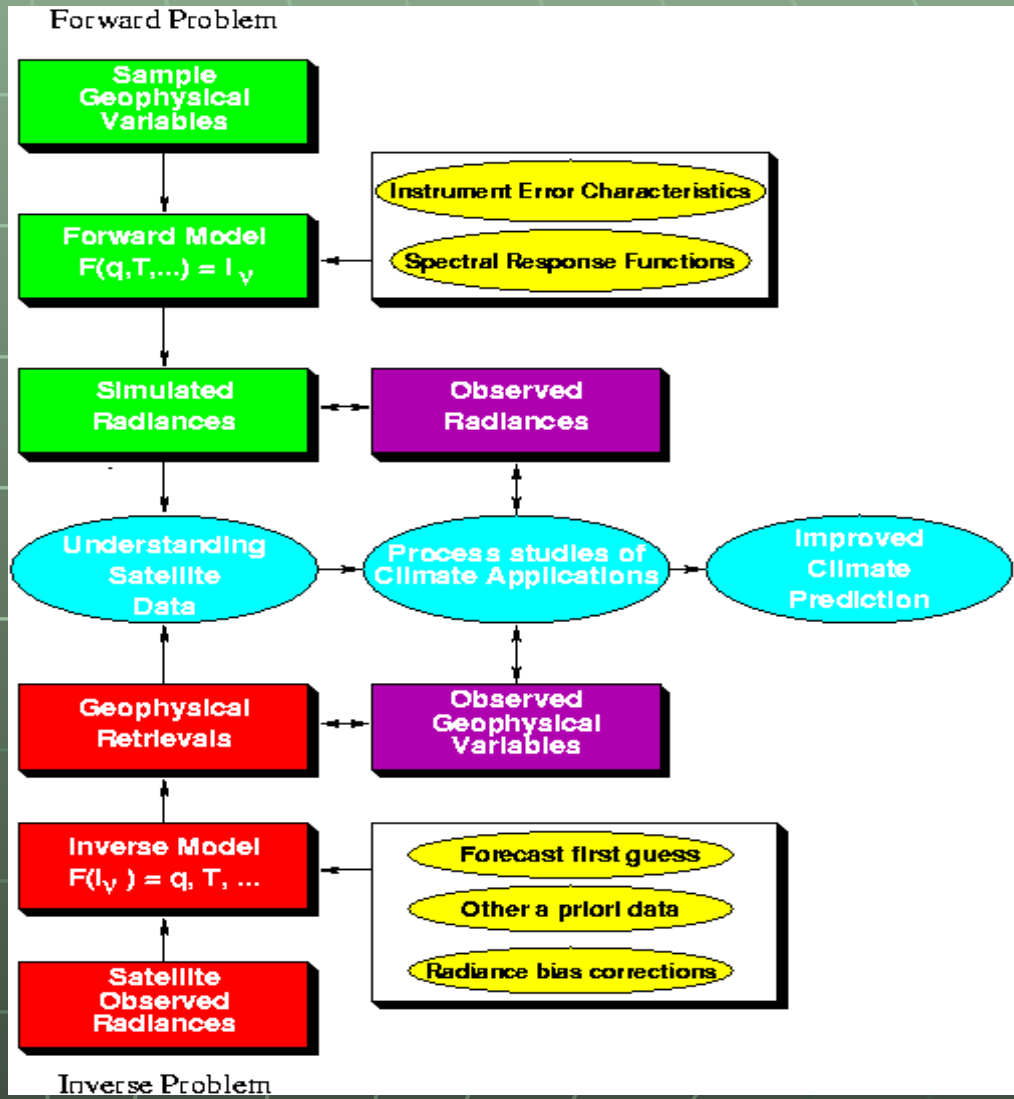
Build and Maintain Highest Quality Information

- Develop and apply common algorithms.
- Apply to Climate Change critical science questions.
- Provide data sets used in National & International Assessments.



"Data Mining"
Extracting Specific Key Information

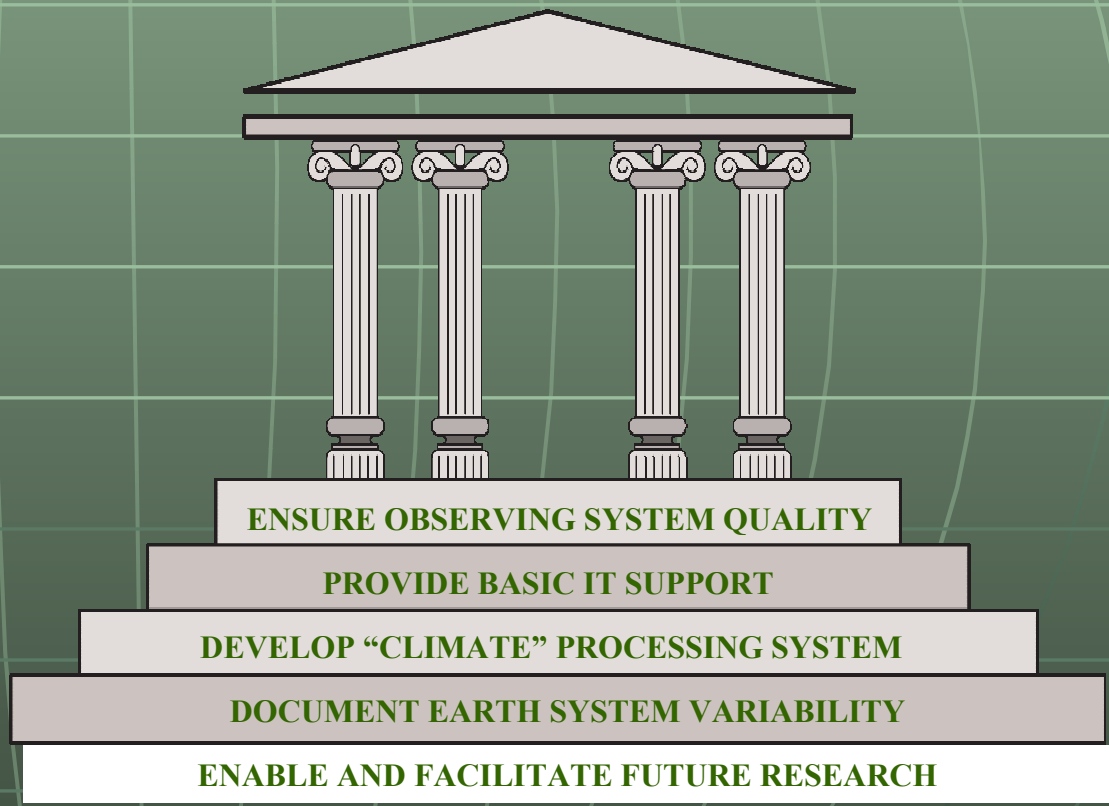
- Information extraction from advanced technology.
- Defining a philosophy of information retrieval for differing user classes and needs.
- Scientific stewardship of observing system requirements.



Scientific Stewardship

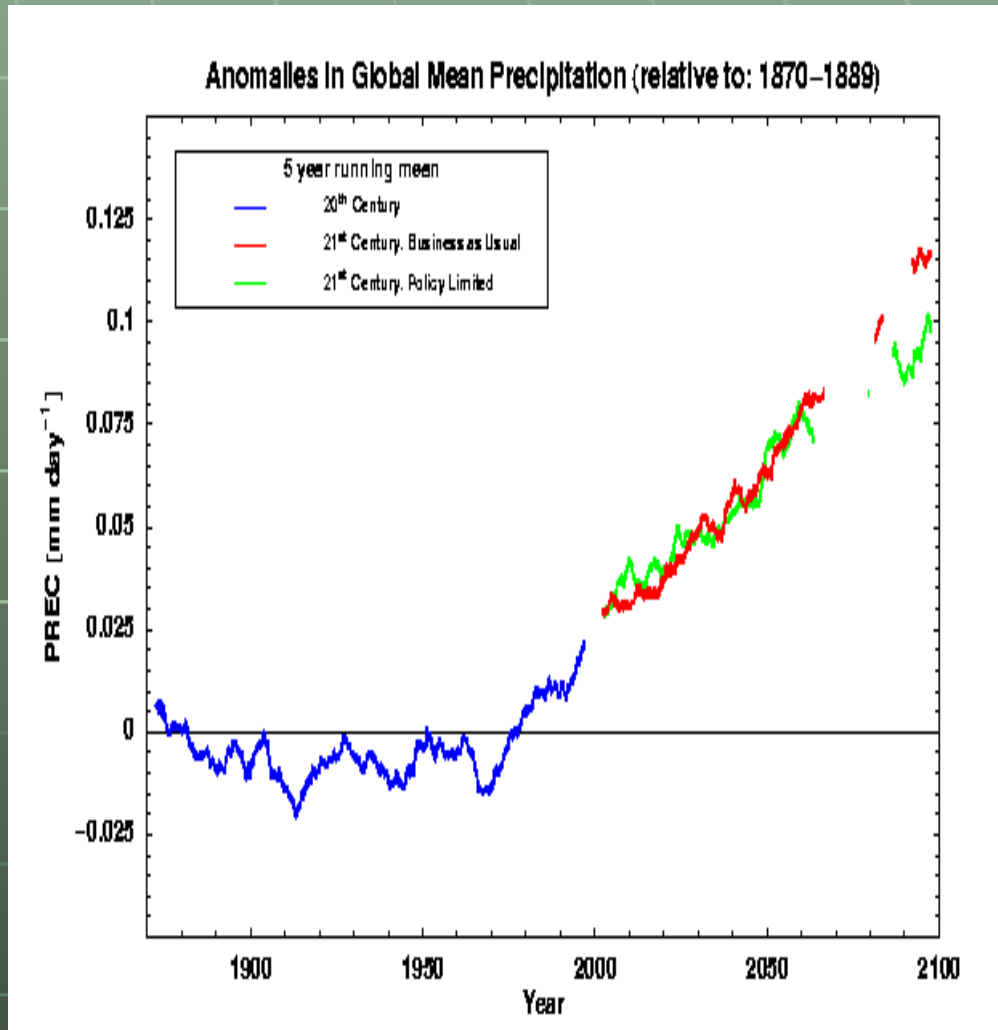
Enable and Facilitate Future Research

- Latest IT tools.
- New environmental change imperative questions.
- Safeguard National Treasure for use by future generations.



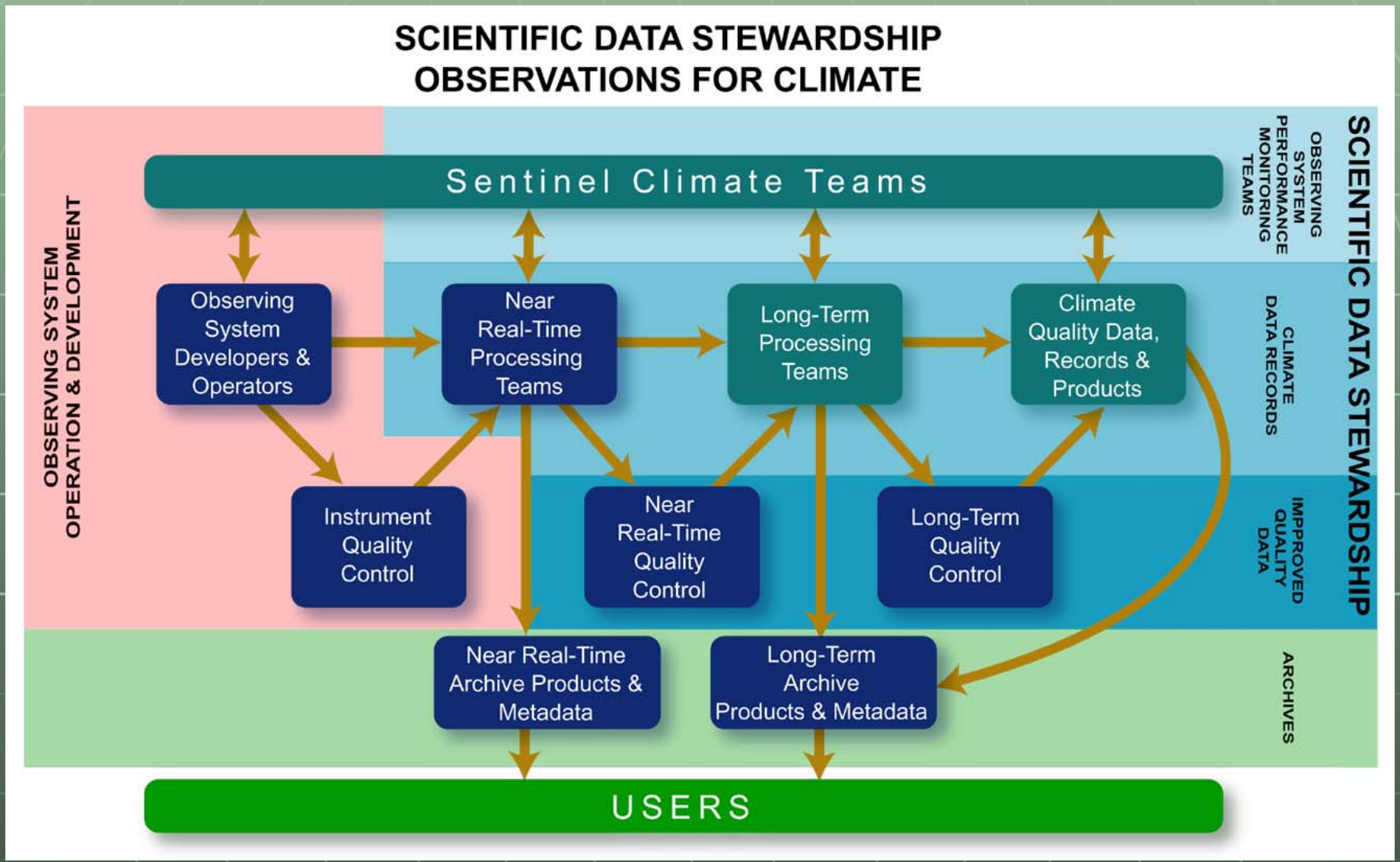
Enable and Facilitate Future Research

- Make data sets easily available through the web.
- Data sets used to update scenarios and assessments.
- Identify and respond to emerging Science Questions.



NOAA Scientific Data Stewardship (SDS)

- To build consistent and high-quality records of environmental observations with associated metadata.
- To partner with the scientific community (and others) through provision of high quality data and services, generation of useful and understandable products, and contributions to scientific communities, including peer reviewed papers.
- To produce comprehensive analyses of environmental change.


SCIENTIFIC DATA STEWARDSHIP
OBSERVATIONS FOR CLIMATE
OBSERVING SYSTEM
OPERATION & DEVELOPMENT
SCIENTIFIC DATA STEWARDSHIP
OBSERVING SYSTEM PERFORMANCE MONITORING TEAMS
CLIMATE DATA RECORDS
IMPROVED QUALITY DATA
ARCHIVES
Sentinel Climate Teams
Observing System Developers & Operators
Near Real-Time Processing Teams
Long-Term Processing Teams
Climate Quality Data, Records & Products
Instrument Quality Control
Near Real-Time Quality Control
Long-Term Quality Control
Near Real-Time Archive Products & Metadata
Long-Term Archive Products & Metadata
USERS

Implementation
Scientific Data Stewardship Groups

- Data Character Group
- Observing Systems Groups: In-situ (ASOS, USCRN. UA, NEXRAD, etc.), Satellite (GOES/POES, NPP, NPOESS, Terra, Aqua, Aura)
- Interdisciplinary Groups – global water, energy, and carbon cycles, long-term consistent and continuous monitoring, etc.
- External Grants and Contracts Program

Implementing SDS
Data Character Group

- Long-term calibration, inter-calibration, and validation of all sensors.
- Collaborates with existing national and international observing system groups.
- Assures customers get highest quality basic data and responds to data quality questions.

Implementing SDS Mission Groups

- Specific to each observing platform/network.
- Forms during observing platform/network design & implementation and then transitions to data character group.
- Partners with science teams.
- Assures competency in specifics of each mission and complete metadata.

Implementing SDS Interdisciplinary Groups

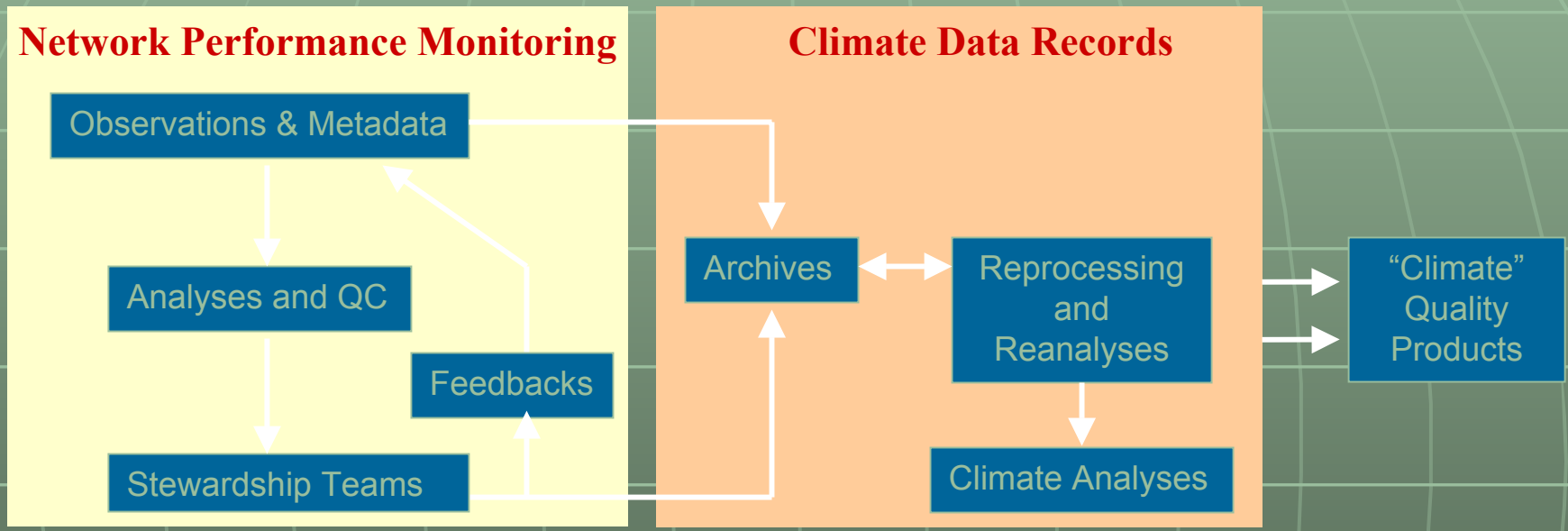
- Address major theme areas: long-term monitoring, water, energy, and carbon cycles, bio & geo chemical cycles, etc.
- Use all instruments, identify key parameters (*Data Mining*) and blend (*Data Fusion*) with all data sources to solve climate change science questions.
- Provide data and information for integrated assessments and options.
- Establish, expand, & reduce as needed.

Implementing SDS
External Grants & Contracts Program

- Works with other SDS groups.
- Directed research using cooperative institutes.
- Use existing and new NOAA grants and contracts program for needed expertise.
- Assure involvement of academia and industry.

NOAA Scientific Data Stewardship

New approach for real time management of climate data



Benefits

- Rapid feedback to observing system
- Data prepared for prediction and analysis
- Model-data synthesis on operational basis
- Simple straight forward data access
- End-to-end accountability of data
  - —Spatial and temporal sampling
  - —Time dependent biases
  - —Metadata
  - —Reprocessing for CDRs
- Enable and facilitate future research
- Safeguard interests of future generations

Summary

- Scientific Stewardship is an Evolving Concept.
- Users want information rather than data.
- Information and products derived from observations are typically more useful to business and industry than the original data.
- Scientists have a critical need for long-time series of quality and continuous historical and recent data to:
  - Assess long-term trends and change
  - Evaluate current variations and trends
  - Predict future conditions and trends

The End Game

Understanding and Knowledge

leading to

Higher Confidence and Improved Forecasts and Predictions

for the

Socioeconomic Benefit of People

and

Environmental Stewardship of Our Planet

Contact Information

Howard J. Diamond, E/CIO

U.S. Global Climate Observing System Program Mgr
NOAA/NESDIS/Office of the CIO
1335 East-West Highway, Room 7214
Silver Spring, MD 20910

E-mail: howard.diamond@noaa.gov

Web Site: http://www.eis.noaa.gov

Voice: +1-301-713-1283 (ext. 229 for voice mail)

Fax: +1-301-713-0819