



# Strategies for the transition and sustainment of climate data records within operations at the National Climatic Data Center

*E. J. Kearns, W. J. Glance, X. Zhao, J. J. Bates,  
J. L. Privette, and H. Semunegus*

AMS Meeting, Phoenix AZ

8 January 2015

US Department of Commerce | NOAA Satellite and Information Service | NOAA's National Climatic Data Center





# Outline

- What is NOAA's operational Climate Data Record (CDR) Program at NCDC
- Evolution in program output and motivations
- Focus on use-inspired products and services
- Update on recent progress and sustainment of operational CDRs

# CDR Program Responding to Rising Demand for Climate Information



Sustainability of Marine Ecosystems



Coasts and Climate Resilience



Climate Impacts on Water Resources



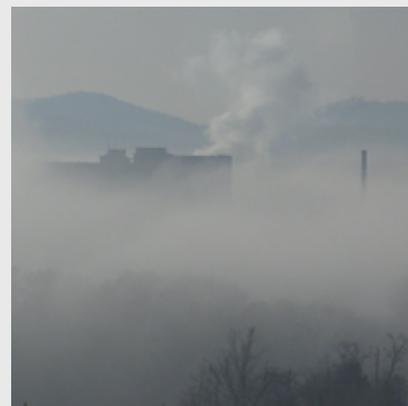
Changes in Extremes of Weather and Climate



Agriculture



Energy

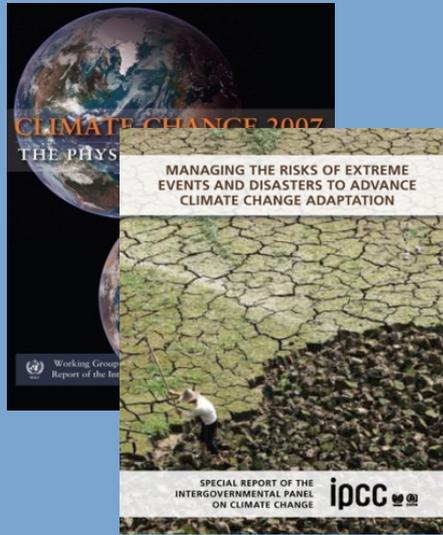


Health



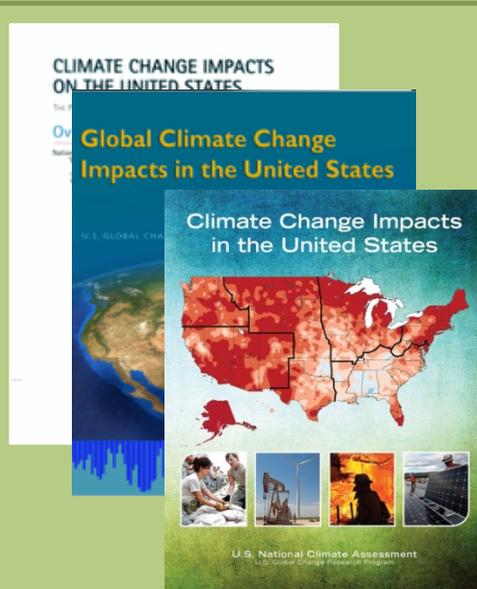
Transportation

# Enabling Assessments of the Earth's Climate: International, National, Annual Assessments



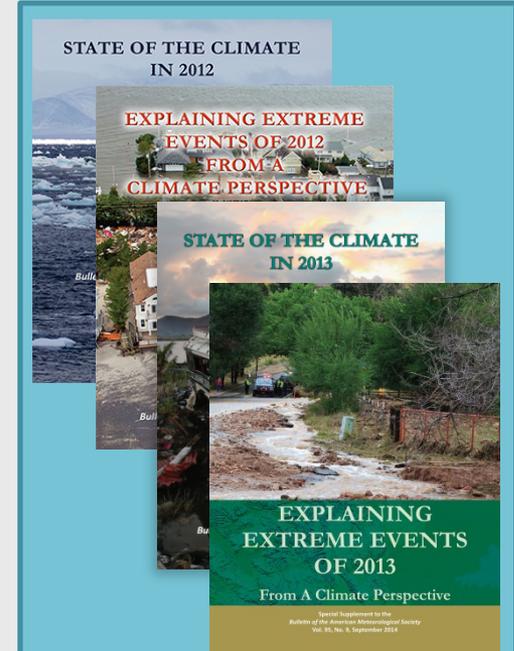
## International Assessments

- 3 NCDC lead authors and review editors on Fourth Assessment Report
- 2 NCDC lead authors on Fifth Assessment Report
- 2 NCDC lead authors on Special Report on Extremes



## National Assessments

- NCDC provides leadership for all National Climate Assessments
- NCDC hosts National Assessment's Technical Support Unit



## Annual Assessments

- NCDC coordinates 378 authors from 48 countries
- Covered by all major news networks, briefed to Hill staff



# What Are CDRs?

- “A **Climate Data Record (CDR)** is a time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change” (US National Academy of Sciences, 2004)
- Relaxed definition: “...sufficient...to have societal value”?
  - **Fundamental CDR (FCDR)**: Calibrated observations for a family of sensors together with the ancillary data used to calibrate them (e.g., Brightness Temperatures)
  - **Thematic CDR (TCDR)**: Geophysical variables derived from FCDRs; may be generated by blending satellite observations, in-situ data, and model output (e.g., Sea Surface Temperature)

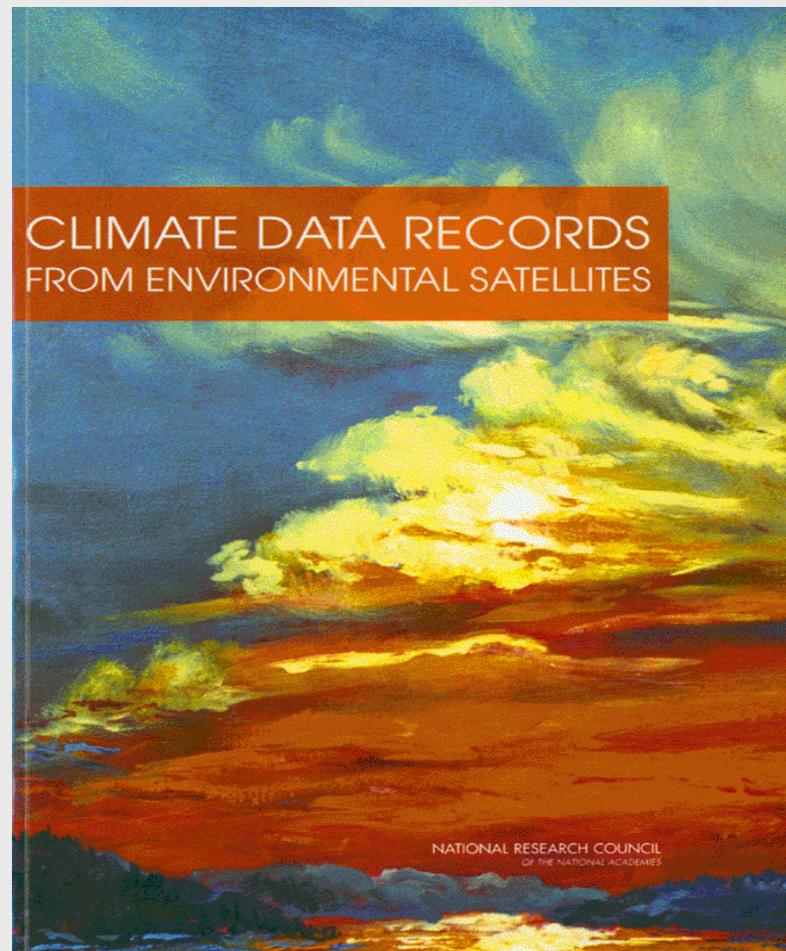


# Why Are NOAA CDRs Important?

- NOAA's satellite CDRs comprise its longest record of global operational satellite measurements. By applying knowledge gathered over time about instruments' performance and sensor characteristics, the data are reprocessed to create consistent and homogenized long-term records.
- NOAA CDRs are sustained in an operational environment, which is critical for supporting decision-making in a changing climate, and thus for the world's resilience to climate changes and variability.

# NOAA CDRP is Well-Grounded in Science and External Expert Guidance

- National Research Council (NRC) of US National Academy of Sciences (NAS) (2004, 2008)
- Global Change Research Program (CCSP, 2006)
- WMO/Global Climate Observing System (GCOS, 2003)
- US EOP/Office of Science and Technology (OSTP), NOAA/NESDIS guidance





# NOAA CDR Program Evolution

1. NOAA's CDR Program began in 2009 at the National Climatic Data Center (NCDC)
  - Selections based on **best science**, most mature algorithms.
  - Focused on **Essential Climate Variables (ECVs)**, with initial priority on **FCDRs** from legacy operational sensors and critical **TCDRs** identified by international and national experts.
  - Competitive **grants** were awarded to capture leading knowledge from the research community.
  - Annual progress meetings ensured NASA, USGS, and international (EUMETSAT) coordination for their related efforts.



# NOAA CDR Program Evolution

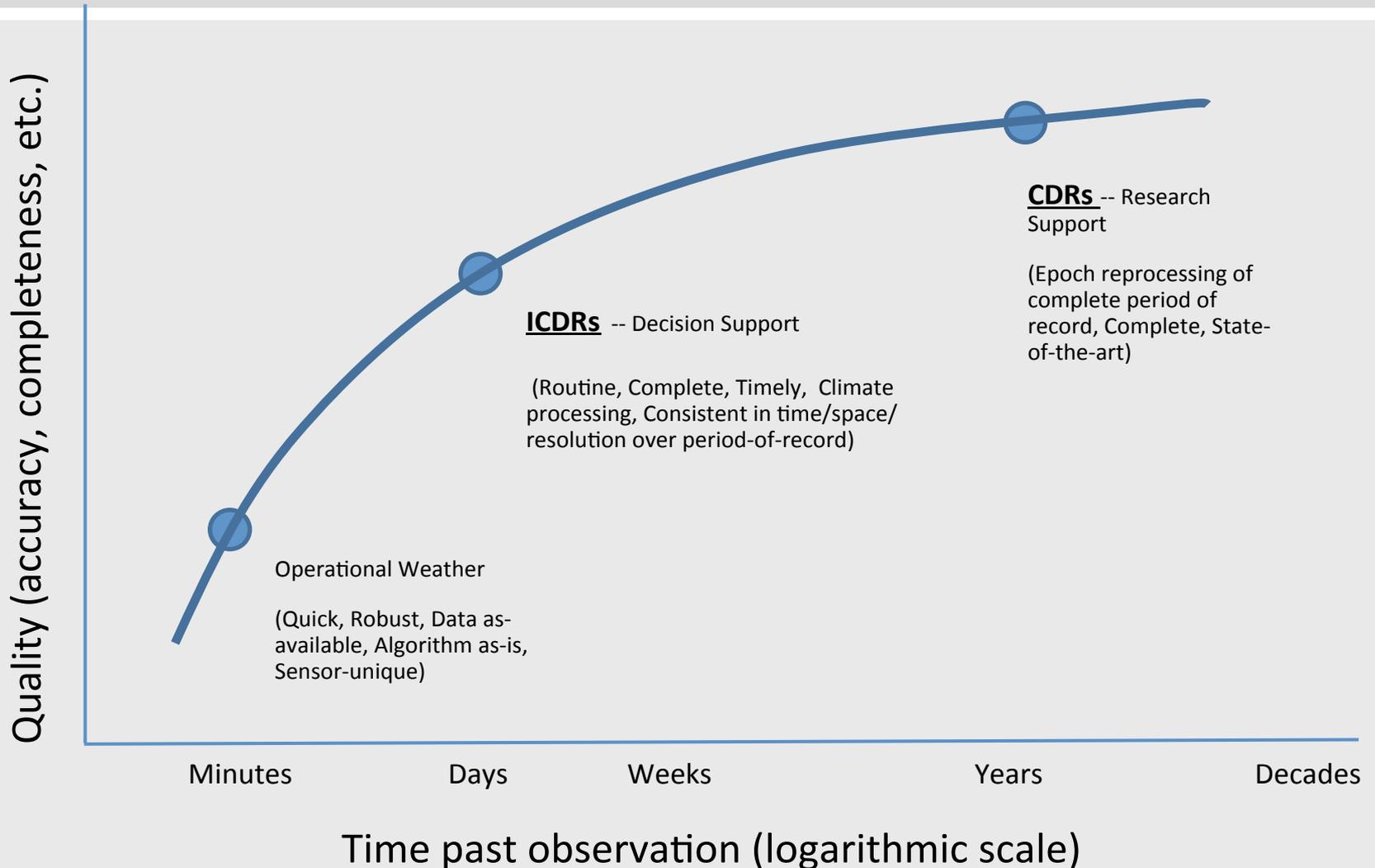
2. CDR Program pivoted in 2012 towards “Use-Inspired” CDRs per NOAA/NCDC guidance
  - Selections based on **best use**, widest applications for societal benefits
    - Grants program concluded.
  - Focused on **engagement with industry and public** to identify their needs and prioritize the CDRs capable of meeting those needs.
  - Refined processes, policies, and procedures for transition from research to operations, and for long-term information preservation.
  - Improved cost estimates for research-to-operations (R2O) transition of CDRs.
  - Some product sustainment and R2O activities are **contracted out**, with project, configuration and risk management practices being applied.



# NOAA CDRP Evolution – 3rd Phase

3. CDRP now (2014+) focusing on implementation of CDRs in practical information products
  - Incorporating ***in situ* and blended products** into the CDR Program to secure those products for the future as well as the satellite data CDRs.
  - Jump-starting **development of applications and Climate Information Products (CIRs)** that target users' needs and requirements, and applying stewardship processes to these applications and CIRs.
  - Support of “Interim” or **ICDRs** & CIRs that project the best known algorithms to near-realtime data flows and products.
  - **Bringing highly-valued but lower-maturity products to a more mature state. E.g. NEXRAD Reprocessing for hi-res precip CDRs.**

# CDRs Supporting Decision Support Applications

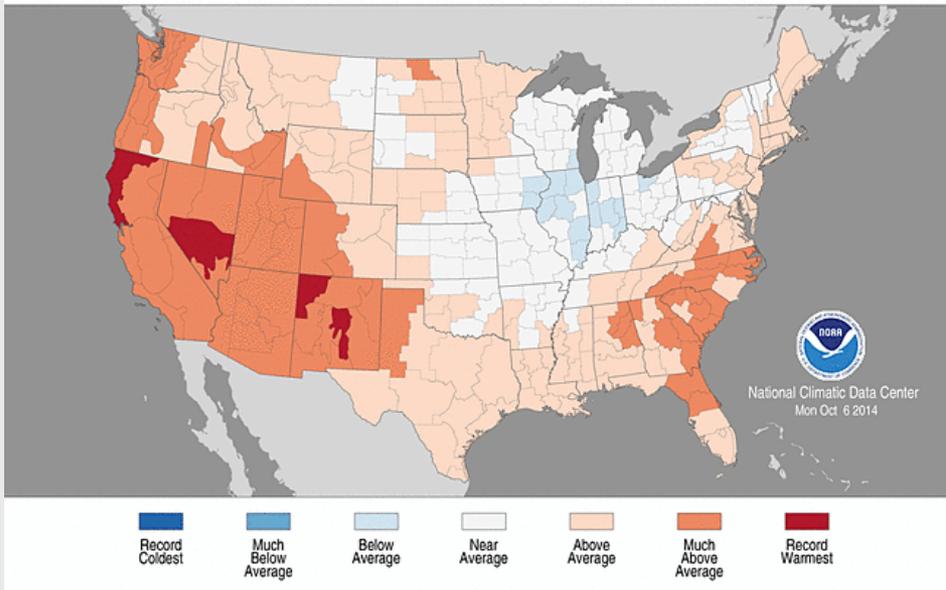


# Decision-makers Increasingly Need Most Salient Points

## Divisional Minimum Temperature Ranks

September 2014

Period: 1895-2014



Can NOAA's operational CDRs/ICDRs meet the sustained demands of:

- Decision makers in industry
- Public sector managers
- Researchers in Earth Sciences

Yes, by way of a consistent, methodical approach.

- **Aspects of Climate Monitoring (e.g. T)**

- **Average** values ( $13.3^{\circ}\text{C}$ )
- Measures of **difference** ( $+1.8^{\circ}\text{C}$  above normal)
- Measures of **unusualness** (much above average = top 10%)
- Measures of **trend** (increasing at  $0.5^{\circ}\text{C}$  per century)
- Measures of **impact** (34% of corn reported in poor condition)



# Climate information platform

- Climate information platforms are emerging through partnerships between industry and government.
- CDRP responsibility encompasses the preservation and stewardship of all the algorithms, code, apps involved in a data set – and needed by information platforms.
- While NOAA's CDRP waits for the infrastructure to be available from/with industry partners, it is focusing on:
  - Making the CDR data ready/portable
  - Describing the data well (documentation, metadata)
  - Preserving and describing the algorithms, workflows, and ancillary data (transparency)
  - Identifying and supporting uses (user requirements)

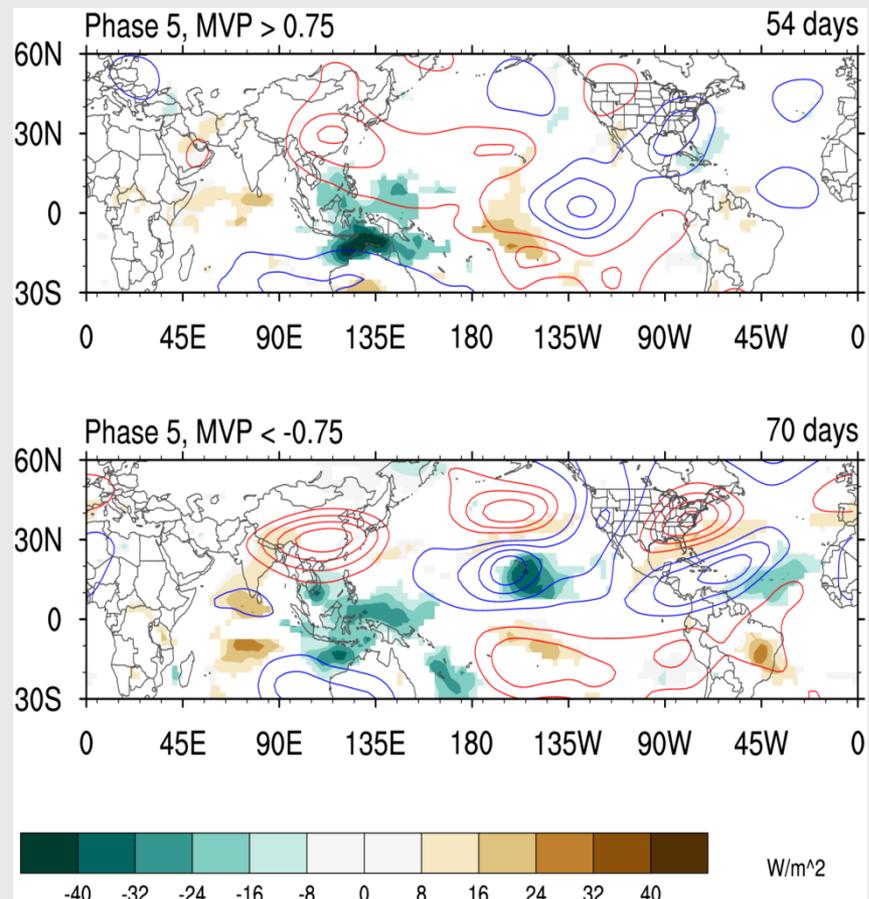
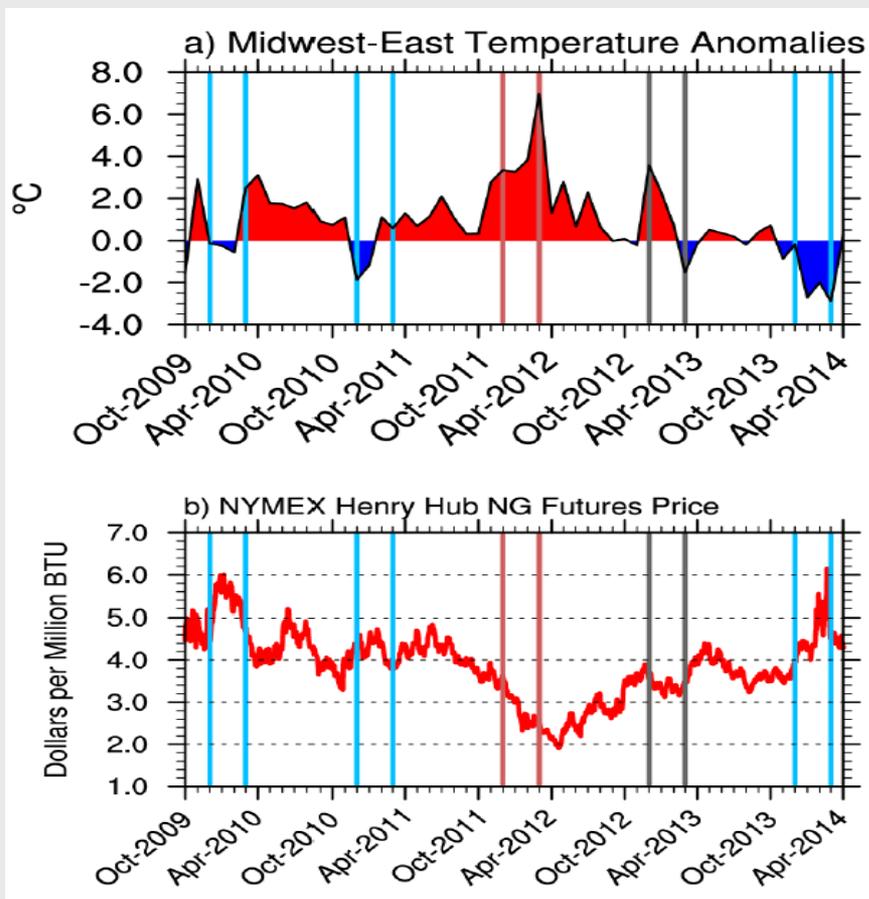
# Sample of NOAA CDR Products

24 CDRs in Ops. as of 2014

<http://www.ncdc.noaa.gov/cdr>

Sample of Operational CDRs			
FCDR	Atmosphere	Ocean	Land
• AVHRR TOA Reflectance	• MW Mean Layer Temperature	• SST (OISST & Pathfinder)	• Surface Reflectance (AVHRR)
• HIRS Brightness Temperature (BT)	• Precipitation (PERSIANN)	• Sea Ice Concentration	• Northern Hemisphere Snow Cover Extent
• SSMI(S) BT	• Cloud (PATMOS-x)		• NDVI (AVHRR)
• VIIRS RCDR	• OLR (HIRS & GridSat)		• LAI/FAPAR (AVHRR)
• MSU/AMSU BT	• Aerosol Optical Thickness (AVHRR)		
• GOES BT (GridSat)			
Research-to-Operation CDRs (work-in-progress)			
FCDR	Atmosphere	Ocean	Land
• Solar Irradiance	• Earth Radiation Budget (ISCCP-ERB)	• Surface Fluxes	• Geo-Surface Reflectance
	• Precipitation (GPCP & CMORPH, NEXRAD)	• Sea Level Height	Snow Concentration
	• Cloud (ISCCP & CERES)		
	• Ozone		

# CIR Provides Commercial Benefit (MVP Index Application)



(Courtesy of Dr. Carl Shreck)

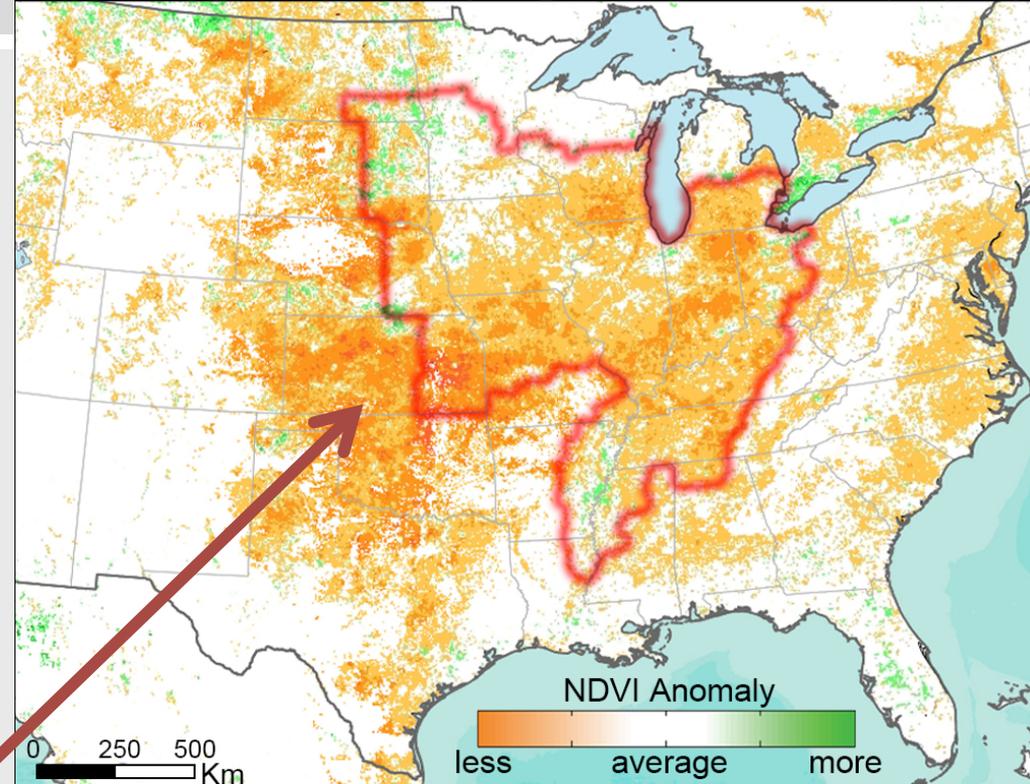
# CDRs Supporting Farming and Agribusiness

## Example: historical context

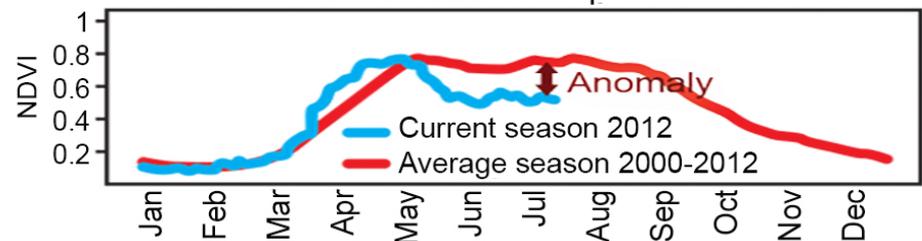
- 5 km resolution, “wall-to-wall” (globally)
- Historical record from 1981- to current
- Collateral products
  - Surface Reflectance
  - Leaf Area Index (LAI)
  - FPAR (photosynthetically active radiation)

Primary U.S. corn and soybean region

2012 drought depicted by Vegetation Index CDR (July 17)



NDVI time series - example in Kansas





# Some Future Challenges

- Publish CDRs as **Linked Data** to promote wider discovery and use.
- **Integrate traceability and integrity mechanisms** into the original data and algorithms available from the Government Archive (e.g., IT security practices).
- Enable or encourage **interoperability** of CDRs with other data types (e.g., health, business, agriculture) and multiple CDRs applied to targeted societal issues (e.g. SSMI and OLR CDRs for Atmospheric River probability studies for Western States precip applications).
- Extend the operational CDRs, which were produced using legacy NOAA operational satellite observations, seamlessly using **new NOAA satellite observations** from Suomi-NPP and the JPSS and GOES-R series.
- **Transfer the capabilities of CDR development, production, preservation, and dissemination** to an emerging cloud-based climate information platform.



# Summary

- The NOAA CDR Program is **well-grounded in science**, is reaching out to address **users' needs**, and continues to improve **open & transparent** stewardship practices for satellite data, non-satellite data and blended products.
  - Includes CIRs, and interim CDR products.
- The NOAA CDR Program at NCDC is **now sustaining 24 satellite data CDRs in operations**, and is preparing CDR data, algorithms, workflows, and documentation for future deployment to a **climate information platform**.
  - CDRs include **not just the data, but the algorithms, workflows, and documentation as well**.
  - With the emerging climate information platform and services from U.S. industry, the Government's role in **Data Stewardship will be more important** than ever.

<http://www.ncdc.noaa.gov/cdr>



Thank you!

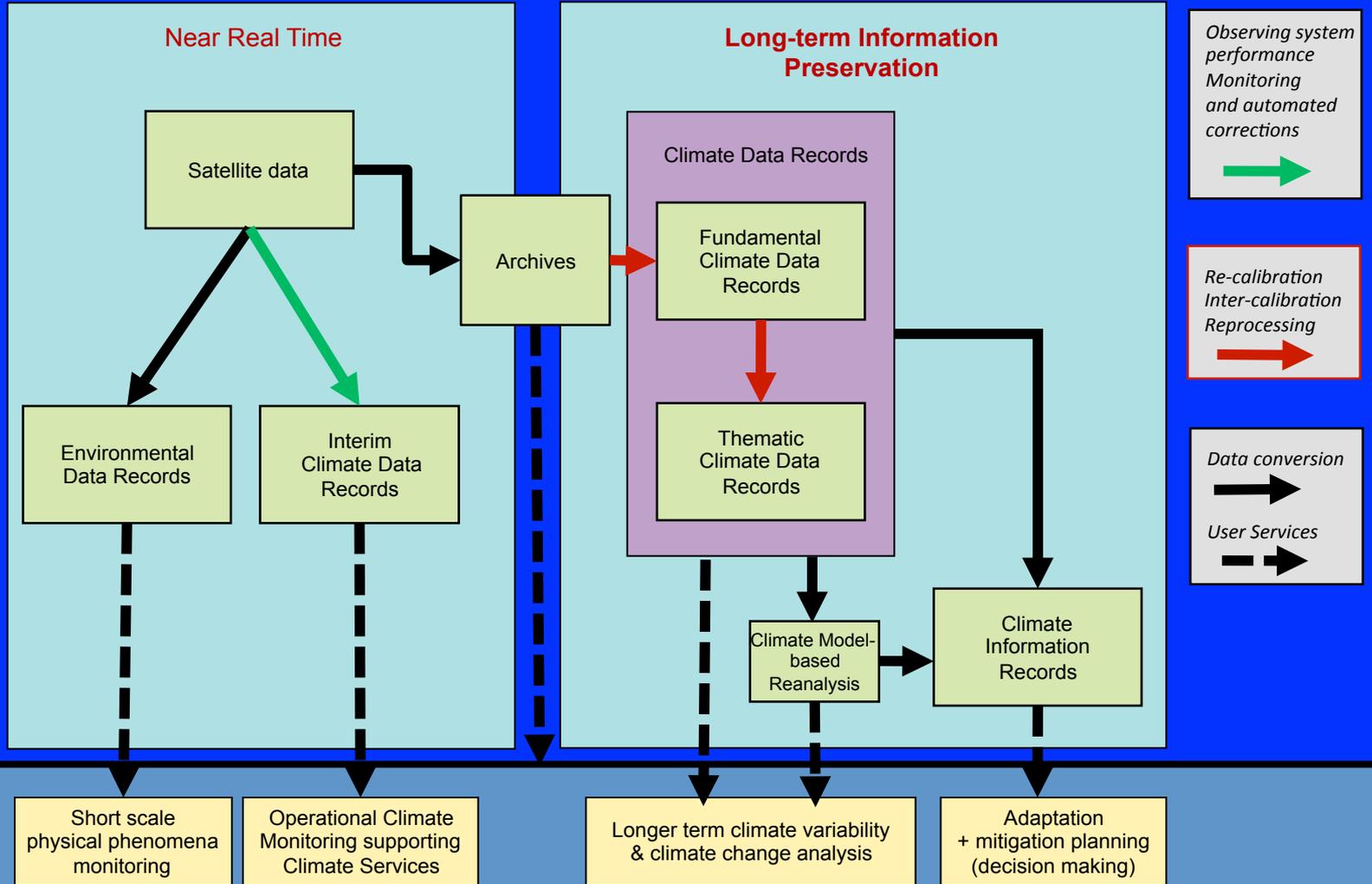
Questions?



# Backup Sides

# NOAA CDRs Sustain Climate Information

Satellite & In Situ Observations



# CDRs Supporting Resource Management

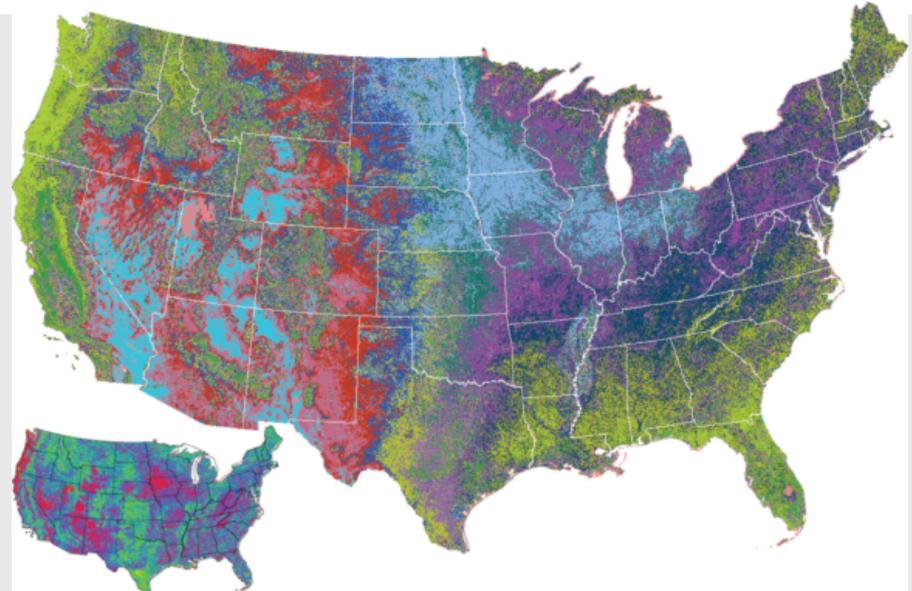
## Example: Forest Change Detection

### Project Goals

- USDA and NASA are developing a capability to assess forest health using a satellite-derived Vegetation Index

### Opportunity

- The CDR Program will leverage NASA and USDA research and transition to operations.



Expanding ranges of forest scavengers linked to climate trends



Gypsy moth larvae

1/8/2015



Pine Beetles

### Output and Impact

- CDR Program produces a 30 year times series of Vegetation Index suitable for identifying forest health trends, leading to better forest and resource management.

### Partners

- USDA Southern Research Station, NASA and NOAA

AMS 2015

# Three Phase Process of NOAA CDRs

## 1. ID

- **Initial Development (ID):** Through grant and contract, PIs develop algorithm, source code, dataset, metadata, and documentation
- PI brings the product to at least Maturity Level-4

## 2. IOC

- **Initial Operational Capability (IOC):** The dataset, metadata, source code and documentation are quality checked, archived and made openly and transparently available for public access.
- Original PI provides operational support and maintenance/updates

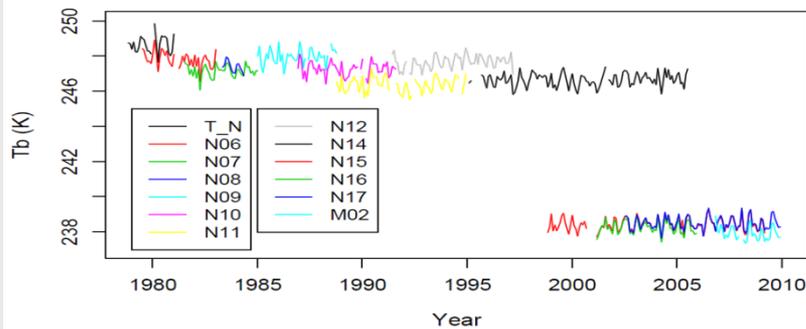
## 3. FOC

- **Full Operational Capability (FOC):** CDR is systematically and routinely generated by NOAA using codes and systems that conform to the NOAA CDR Program's IT security, coding and documentation standards
- CDR operational support and maintenance/updates can be accomplished independent of the original PI [Maturity Level-6]

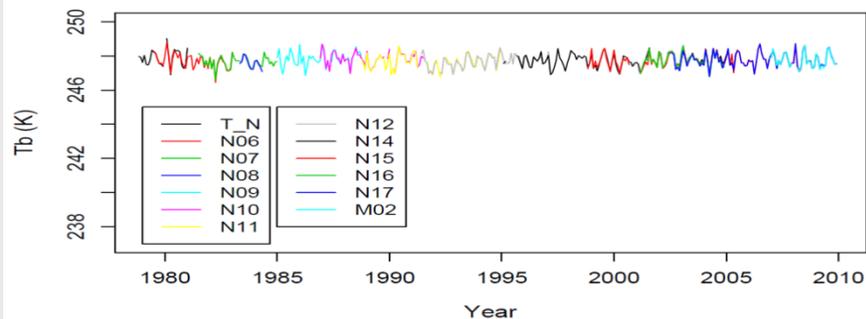
# Why Recommend NOAA CDRs to Users?

(High Quality Data: Inter-calibration and Homogenization Reduce Artifacts Imparted by Observing Systems, Facilitating Meaningful Comparisons in Space and Time)

HIRS BT Timeseries, before inter-calibration



HIRS BT Timeseries, after inter-calibration



**Operational weather products** are produced rapidly to potentially save life and property



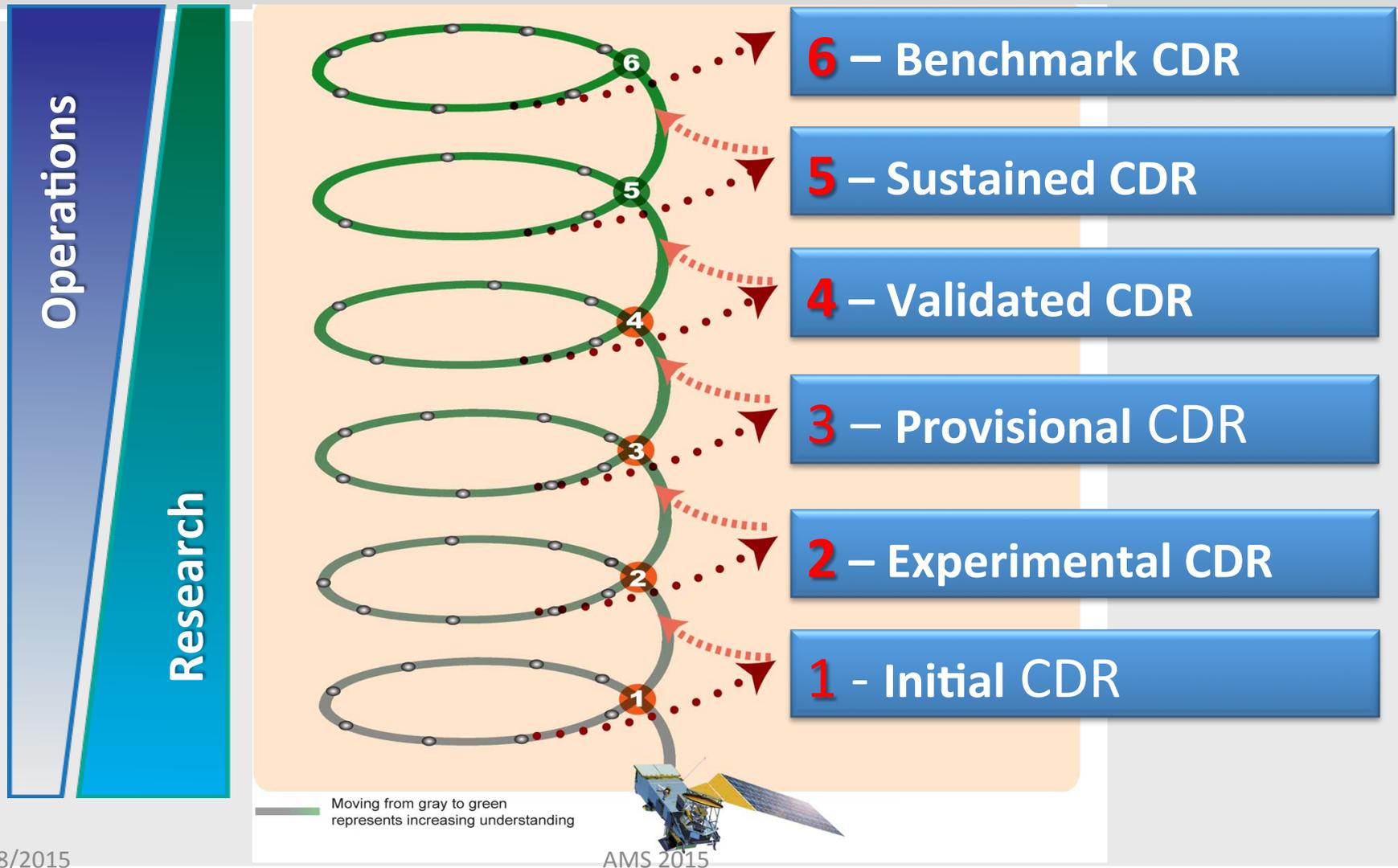
**Climate Data Records (CDRs)** describe climate through rigorous cross-calibration and reprocessing with advanced algorithms, ancillary data and evolved instrument understanding.



# “Maturity Matrix” Defines CDR Product Readiness

Level	Sensor Use	Code Stability	Metadata & QA	Documentation	Validation	Public Release	Science & Applications
1	Research Mission	Significant changes likely	Incomplete	Draft ATBD	Minimal	Limited data availability to develop familiarity	Little or none
2	Research Mission	Some changes expected	Research grade (extensive)	ATBD Version 1+	Uncertainty estimated for select locations/times	Data available but of unknown accuracy; caveats required for use.	Limited or ongoing
3	Research Missions	Minimal changes expected	Research grade (extensive); Meets international standards	Public ATBD; Peer-reviewed algorithm and product descriptions	Uncertainty estimated over widely distribute times/location by multiple investigators; Differences understood.	Data available but of unknown accuracy; caveats required for use.	Provisionally used in applications and assessments demonstrating positive value.
4 (IOC)	Operational Mission	Minimal changes expected	Stable, Allows provenance tracking and reproducibility; Meets international standards	Public ATBD; Draft Operational Algorithm Description (OAD); Peer-reviewed algorithm and product descriptions	Uncertainty estimated over widely distribute times/location by multiple investigators; Differences understood.	Source code released; Data available but of unknown accuracy; caveats required for use.	Provisionally used in applications and assessments demonstrating positive value.
5	All relevant research and operational missions; unified and coherent record demonstrated across different sensors	Stable and reproducible	Stable, Allows provenance tracking and reproducibility; Meeting international standards	Public ATBD, Operational Algorithm Description (OAD) and Validation Plan; Peer-reviewed algorithm, product and validation articles	Consistent uncertainties estimated over most environmental conditions by multiple investigators	Source code portable and released; Multi-mission record is publicly available with associated uncertainty estimate	Used in various published applications and assessments by different investigators
6 (FOC)	All relevant research and operational missions; unified and coherent record over complete series; record is considered scientifically irrefutable following extensive scrutiny	Stable and reproducible; homogeneous and published error budget	Stable, Allows provenance tracking and reproducibility; Meeting 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	Source code portable and released; Multi-mission record is publicly available from Long-Term archive	Used in various published applications and assessments by different investigators

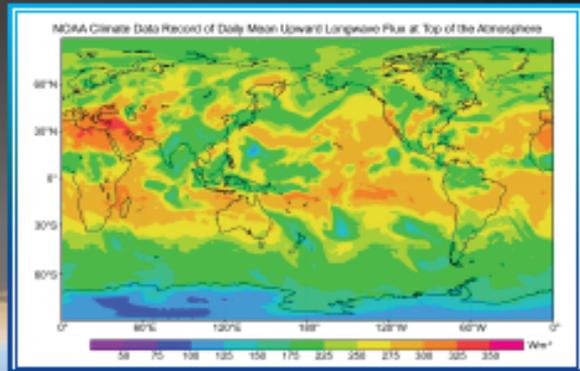
# 6-Level Maturity Model of NOAA CDRs



# NOAA's Climate Data Record (CDR) Program

## OUTGOING LONGWAVE RADIATION - DAILY

# OLR - Daily



### OLR - DAILY CLIMATE DATA RECORD SPECIFICATIONS

- Global Coverage
- 1.0x1.0 Degree Equal-Angle Grid
- Daily Mean Product
- 1979 – Present
- Updated Quarterly
- Interim CDR Available within 48 Hours of Observation

### INPUTS TO THE OLR - DAILY CLIMATE DATA RECORD

- High-resolution Infrared Radiation Sounder (HIRS) Level-1b Data
- GridSat Geostationary Imager Brightness Temperatures
- GSIP (GOES Surface and Insolation Product) for Interim CDR
- OLR Regression Coefficients
- Calibration Prediction Coefficients
- Inter-satellite Bias Corrections

### SOME USES OF THE OLR - DAILY CLIMATE DATA RECORD

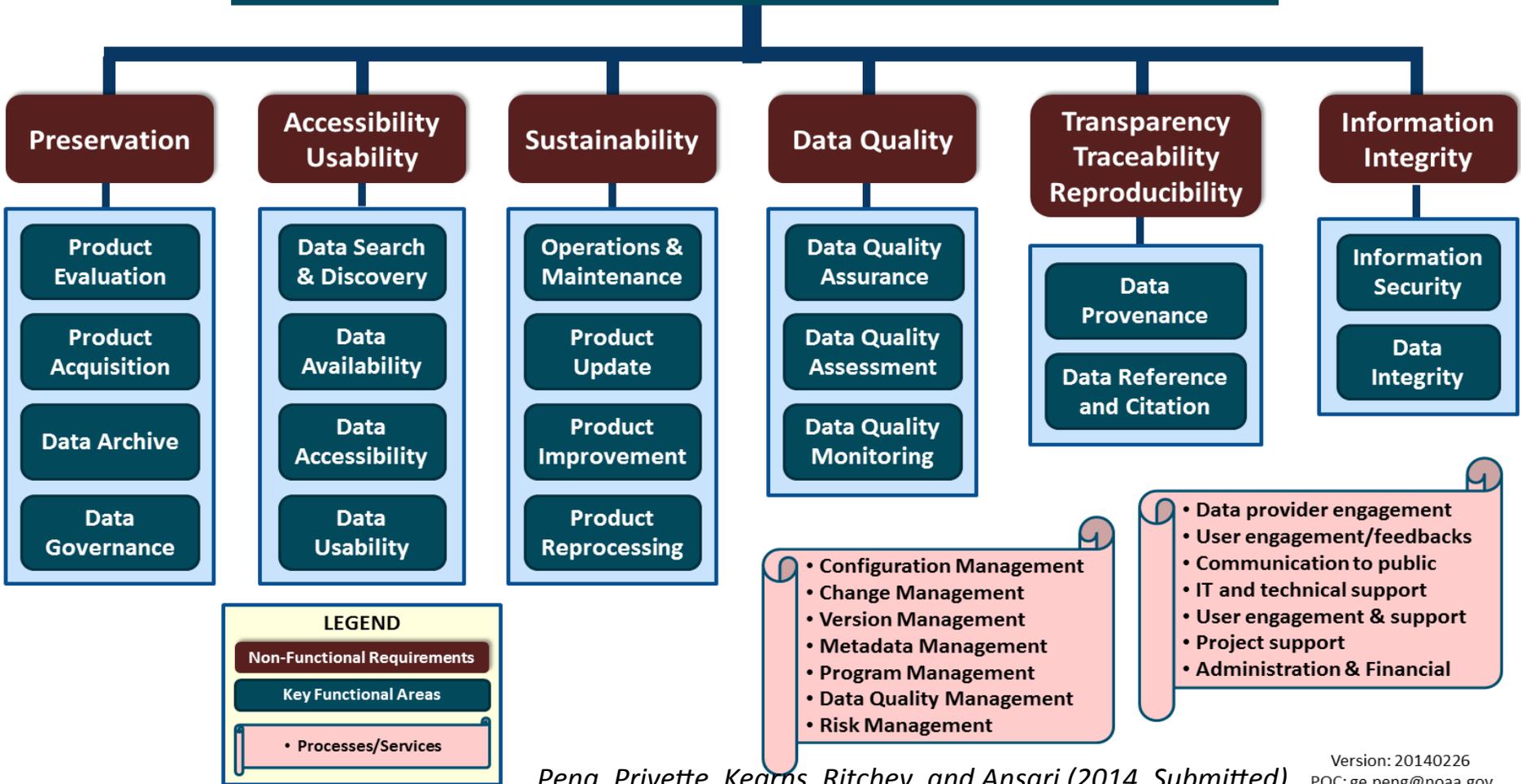
- Input into Radiation Budget Studies
- Verifying Numerical Models
- Studying Short-Term and Long-Term Climate Variability
- Preparing Diagnostics and Forecasts of the MJO and Tropical Waves
- Analyzing and Predicting Global Precipitation Patterns
- Predicting Global Tropical Cyclone Activity

**OLR - DAILY CLIMATE DATA RECORD**  
<http://www.ncdc.noaa.gov/cdr/operationalcdrs.html>

**CLIMATE DATA RECORD PROGRAM INFORMATION**  
<http://www.ncdc.noaa.gov/cdr/index.html>



# Long-Term Scientific Data Stewardship



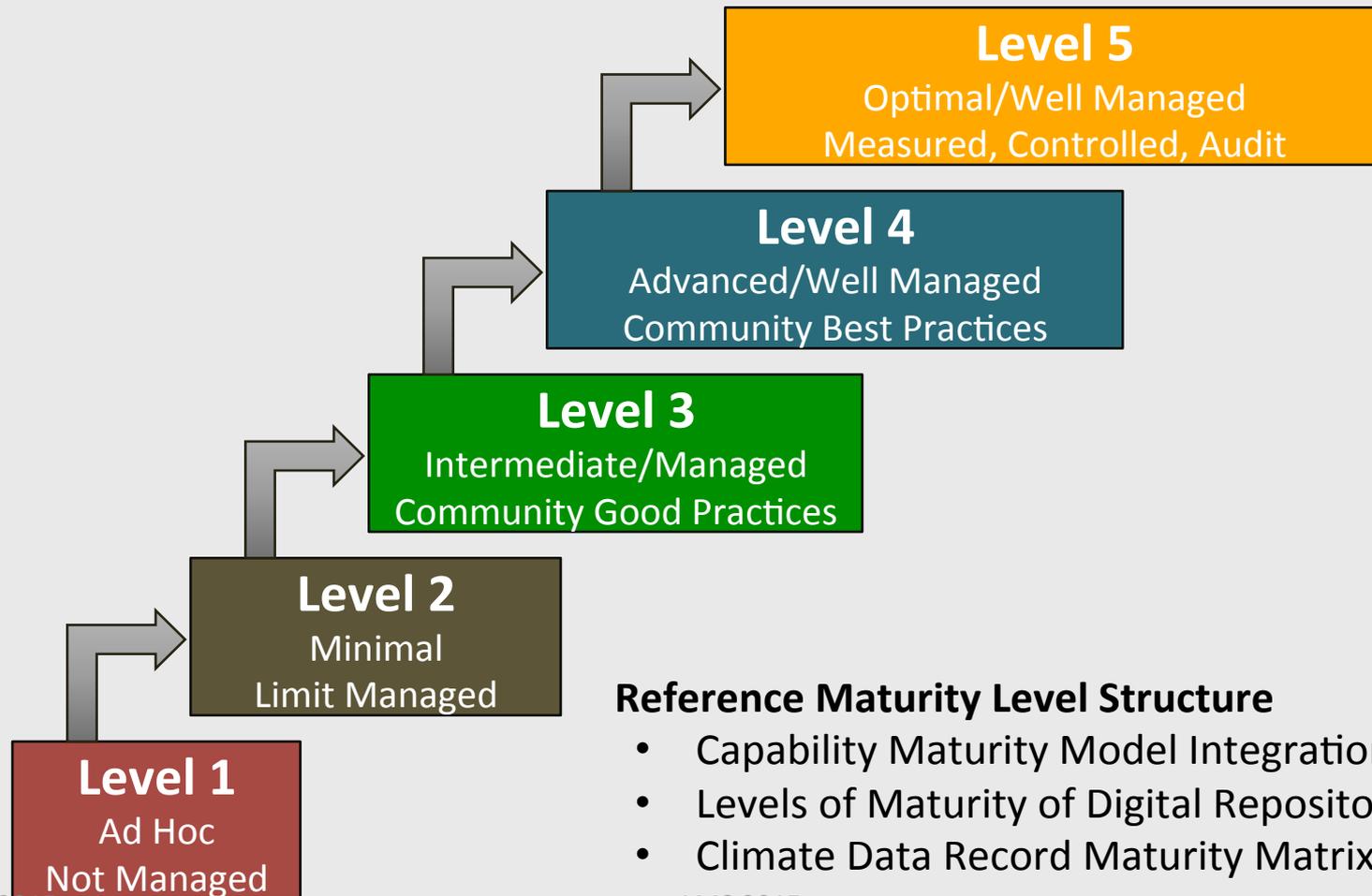
Peng, Privette, Kearns, Ritchey, and Ansari (2014, Submitted)

Version: 20140226  
POC: ge.peng@noaa.gov

## Data Stewardship

All activities that preserve and improve the information content, accessibility, and usability of data and metadata (NRC, 2007) and that ensure or improve the quality and usability of environmental data

# Proposed Maturity Levels Follow the CMMI Structure



## Reference Maturity Level Structure

- Capability Maturity Model Integration (CMMI)
- Levels of Maturity of Digital Repositories
- Climate Data Record Maturity Matrix (CDRMM)

# Dataset Name

Maturity Level as of  
mm/dd/yyyy

## Stewardship Maturity Matrix for Digital Environmental Data Products

Maturity Scale	Preservation	Accessibility	Data Integrity	Usability	Production Sustainability	Data Quality Screening	Data Quality Assurance	Data Quality Monitoring/Control	Data Quality Assessment	Transparency /Traceability
<b>Level 1 – Ad Hoc Not Managed</b>	Any storage location Data only	Not publicly available Person-to-person	Unknown or no data integrity check	Extensive product-specific knowledge required No document online	Ad Hoc No obligation or deliverable requirement	Unknown or none	Data quality assurance (DQA) procedure unknown or none	Sampling unknown or spotty Analysis unknown or random in time	Algorithm theoretical basis assessed	Limited product information available Person-to-person
<b>Level 2 - Minimal Managed Limited</b>	Non-designated repository Redundancy Limited archiving metadata	Publicly available Not searchable online	Data ingest integrity verifiable (e.g., checksum technology)	Non-standard data format Limited document (e.g., user's guide) online	Short-term Individual PI's commitment (grant obligations)	Completeness and redundancy check	Ad Hoc and random Procedure not defined and documented	Sampling and analysis are regular in time and space Limited product-specific metric defined & implemented	Level 1 + Research product assessed	Product information available in literature
<b>Level 3 - Intermediate Managed Defined, Partially Implemented</b>	Designated repository/archive Redundancy Conforming to community archiving process and metadata Conforming to limited archiving standards	Available online Limited data server performance Collection/ dataset searchable	Level 2 + Data archive integrity verifiable	Community Standard-based interoperable format & metadata Documentations (e.g., source code, product algorithm document, processing or/and data flow diagram) online	Medium-term Institutional commitment (contractual deliverables with specs and schedule defined)	Level 2 + Consistency check (i.e., spatial, temporal, logical)	Procedure defined and documented and partial implemented	Sampling and analysis are regular, frequent, systematic but not automatic Community metric defined and partially implemented Procedure documented and available online	Level 2 + Operational product assessed	Algorithm Theoretical Basis Document (ATBD) & source code online Dataset configuration managed (CM) Data citation tracked (e.g., utilizing DOI system) Unique object identifier (UOI) assigned (dataset, documentation, source code)
<b>Level 4 - Advanced Managed Well-Defined, Fully Implemented</b>	Level 3 + Conforming to community archiving standards	Level 3 + Enhanced data server performance Granule/file searchable	Level 3 + Data access integrity verifiable Conforming to community data integrity technology standard	Level 3 + Basic capability (e.g., subsetting, aggregating) & data characterization (overall/global, e.g., climatology, error estimates) available online	Long-term Institutional commitment Product improvement process in place	Level 3 + Accuracy check Procedures well documented and available online Limited quality screening metadata	Procedure well documented and available online with master reference data Limited data quality assurance metadata	Level 3 + Anomaly detection procedure well-documented and fully implemented using community metric, automatic, tracked and reported Limited quality monitoring metadata	Level 3 + Quality metadata assessed Limited quality assessment metadata	Level 3 + OAD (Operational Algorithm Description) online (CM + UOI)
<b>Level 5 - Optimal Level 4 + Measured, Controlled, Audit</b>	Level 4 + Archiving process performance controlled, measured, and audited Future archiving process and standard changes planned	Level 4 + Dissemination reports available online Future technology changes planned	Level 4 + Data authenticity verifiable (e.g., data signature technology) Performance of data integrity check monitored and reported	Level 4 + Enhanced online capability (e.g., visualization, multiple data formats) Community metric set of data characterization (regional/cell) online External ranking	Level 4 + National or international commitment Changes for technology planned	Level 4 + Conforming to community quality metadata & standards	Level 4 + Conforming to community quality metadata & standards External review	Level 4 + Cross-validation of temporal & spatial characteristics Consistency check Conforming to community quality metadata & standards Dynamic providers/users feedback in place	Level 4 + Assessment performed on a recurring basis Conforming to community quality metadata & standards External ranking	Level 4 + System information online Complete data provenance

Level 1: Red | Level 2: Black | Level 3: Green | Level 4: Blue | Level 5: Gold

AMS 2015

Version: SDS\_MM\_20140527\_beta\_v1.0  
POC: [Ge.Peng@noaa.gov](mailto:Ge.Peng@noaa.gov); [Jeff.Privette@noaa.gov](mailto:Jeff.Privette@noaa.gov)

# Operational CDRs Will Cover Three Major NOAA Satellite Epochs

1970                      1980                      1990                      2000                      2012                      2020                      2030

POES/GOES/DMSP

NPP

JPSS/JASON-3/GOES-R

Reveal latent climate trend information  
in four decades of heritage operational data

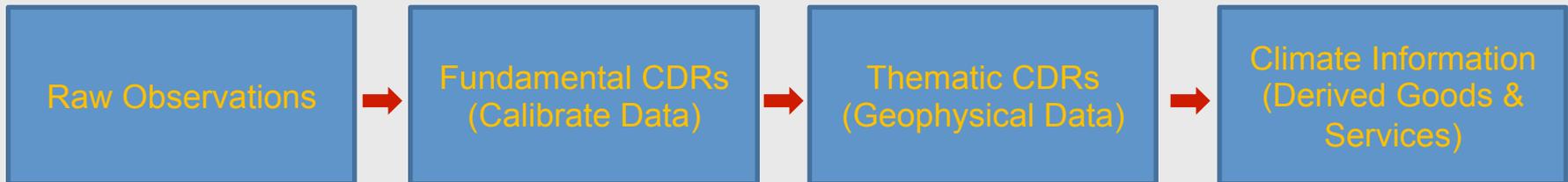
Extend CDRs  
using future  
sensors

Ensure climate quality data from new operational  
system and extend CDRs period of record

# CDRs Supporting Energy Sector

Example: Solar Farm location siting

A collaboration between NOAA, NASA, Academia, the Private Sector and NREL

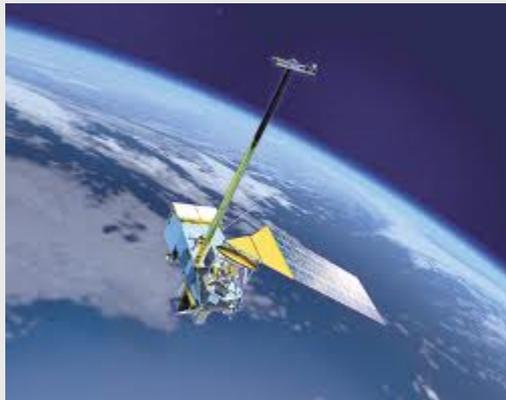


NOAA & NASA  
raw satellite data

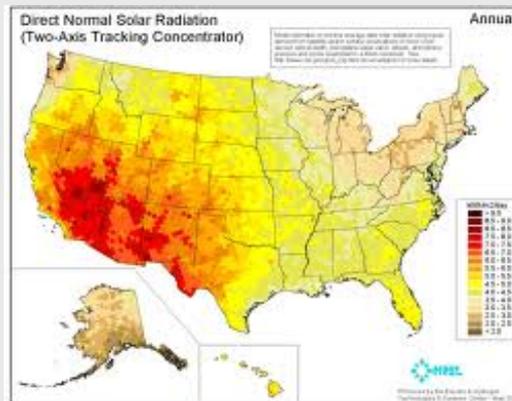
Merged and corrected  
NOAA & NASA  
data provide trustworthy  
multi-decadal  
time series

Standards of the  
renewable  
energy community  
applied

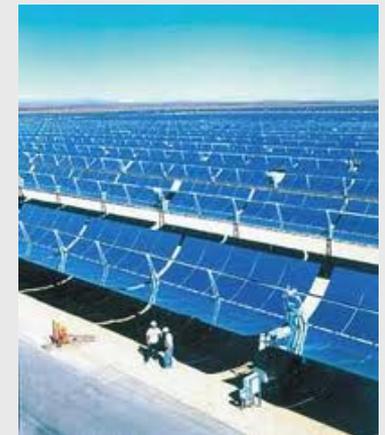
Siting information for  
major solar energy  
farms



1/8/2015



AMS 2015

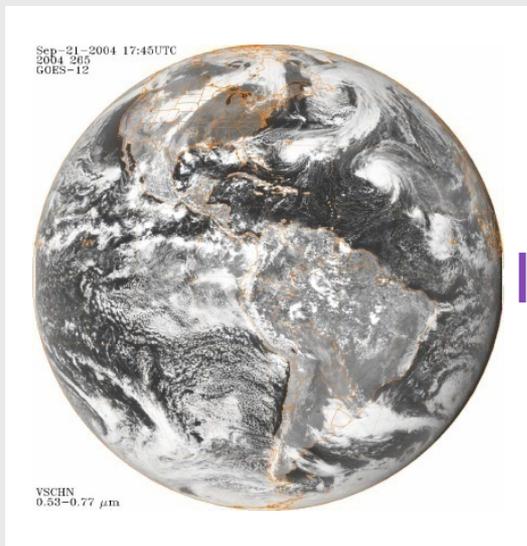


NREL= National Renewable Energy Laboratory

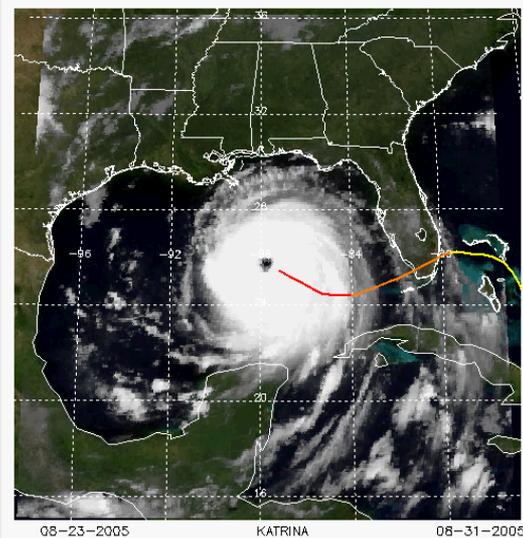
# CDRs Supporting Insurance/Reinsurance

## Example: Hurricane Trends

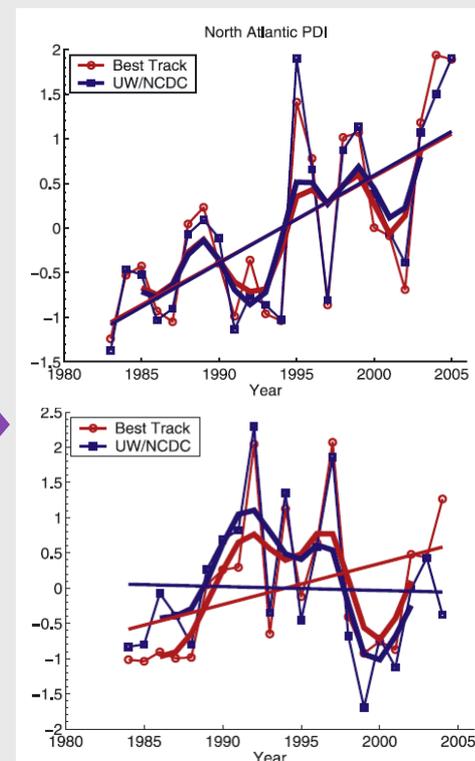
Government provision of data



Transition from government to industry



Decision support information



Hurricane intensity trends (Kossin et al. 2007)

# The Reference Model for an Open Archival Information System (OAIS)

The OAIS Reference Model attempts to comprehensively identify the responsibilities and components of an archival system, including:

- the roles of people and institutions that interact in an archive: Producer, Management, and Consumer
- the digital objects: information package
- the major functions: six higher-level functions – Ingest, Data Management, Archival Storage, Access, Preservation Management, and Administration, representing thirty-three lower-level functions

<http://www.dpworkshop.org/dpm-eng/foundation/oais/index.html>

