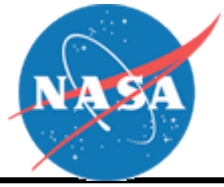
CERES Data Management Activity



Presented to

CERES Science Team

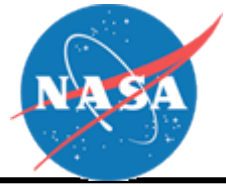
NASA Langley Research Center

November 3, 2009

Jonathan Gleason

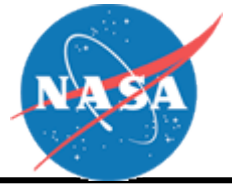
jonathan.l.gleason@nasa.gov

Topics to be Covered



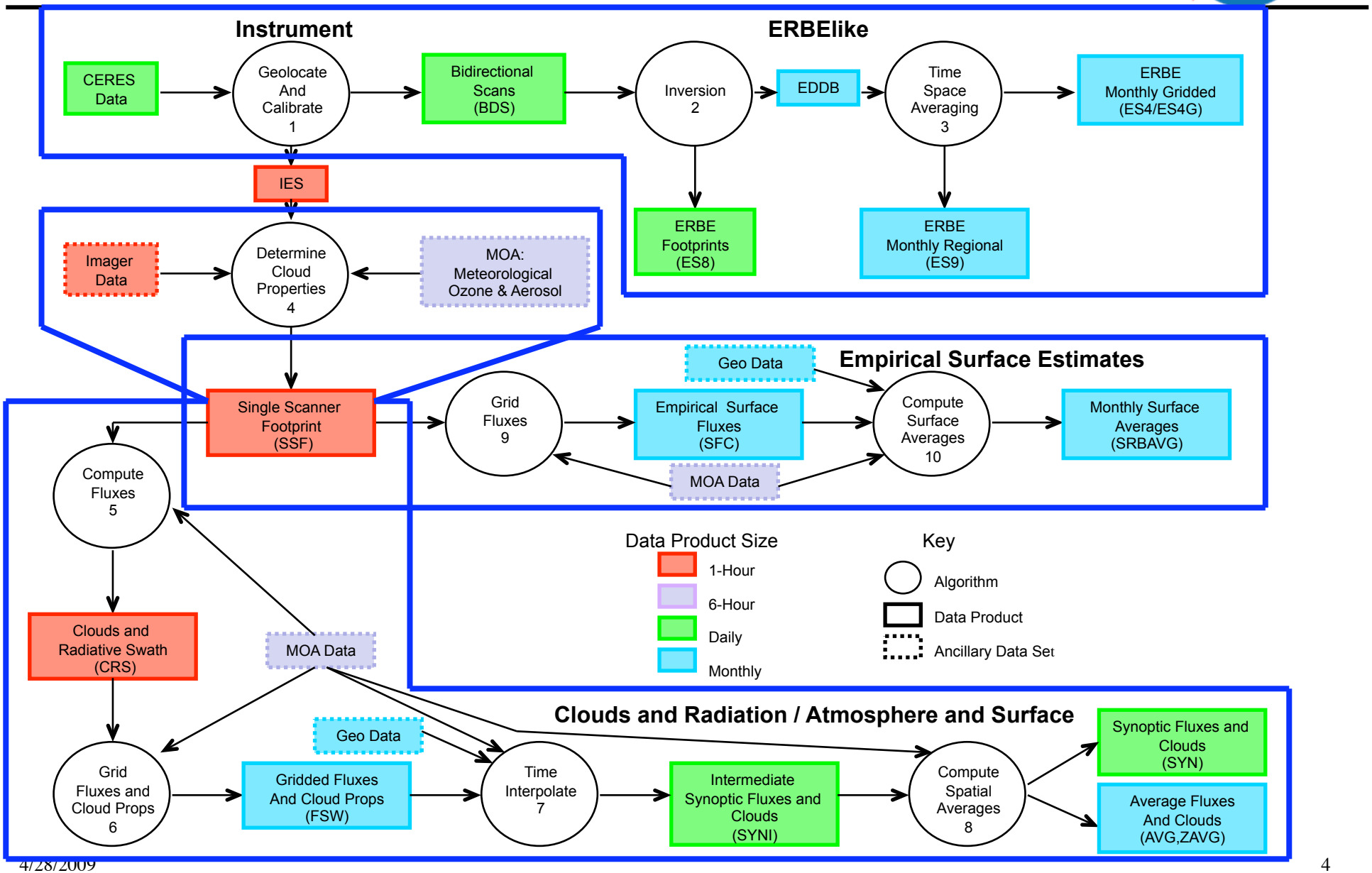
- CERES Background/Overview
- CERES Processing Approach
- NPP as it compares to Terra/Aqua
- Terra/Aqua Edition 2 & Edition 3 status
- Development and Production Platforms
- AMI Migration Status
- Production optimization effort
- DMT Operations and Documentation

CERES Statistics
(from the Terra & Aqua Senior Review)

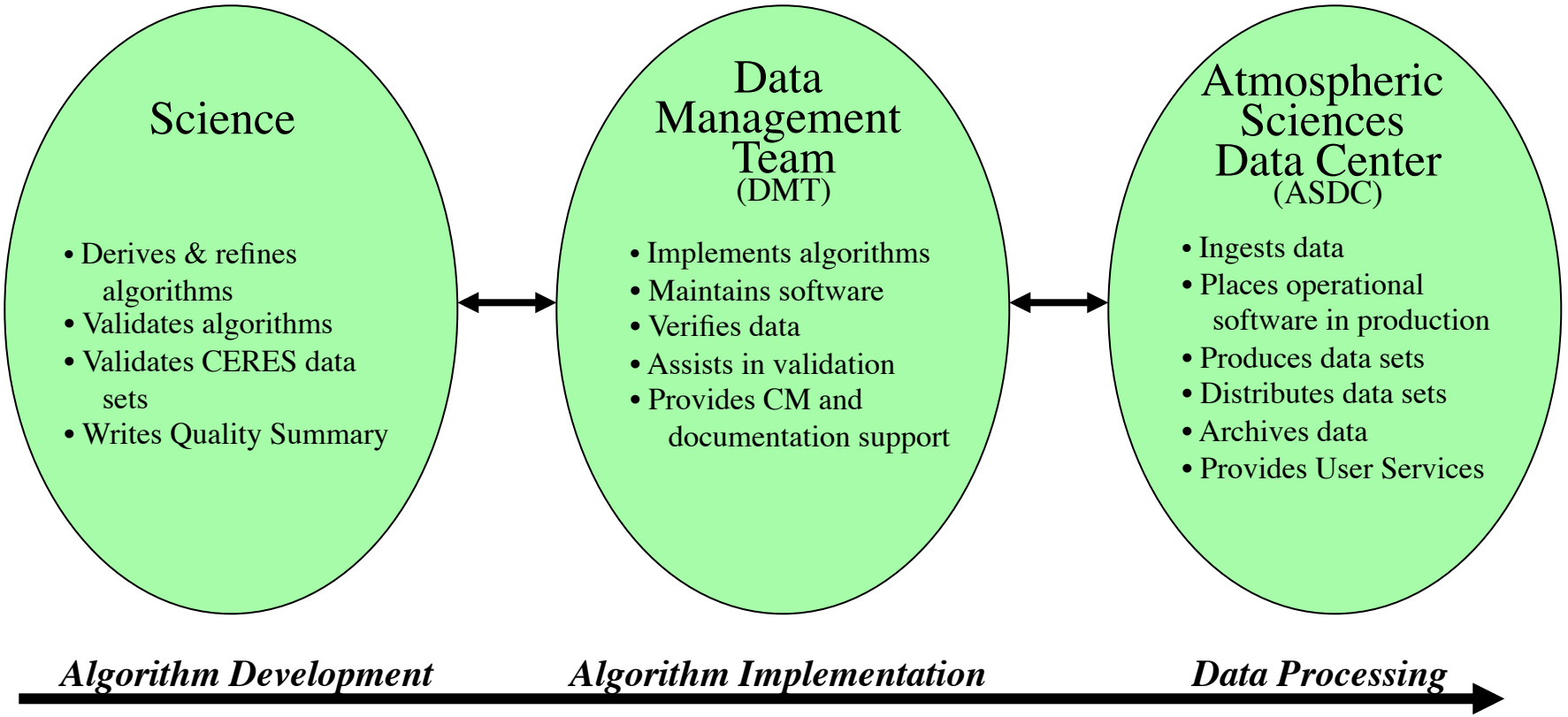


- High level of data fusion
  - 11 Instruments on 7 satellites
- 25 unique input data sources
- 18 CERES data products
- Over 90% of CERES data product data volume involves 2+ instruments
- Individual data products include up to 260 unique parameters
- Approximately 1.7 million lines of QC and validation codes
- Approximately 0.85 million lines of production codes

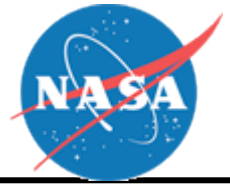
CERES Top Level Data Flow Diagram



CERES Organization



CERES Subsystems



- CERES is made up of 7 Working Groups
  - Instrument
  - ERBElike
  - Clouds
  - Inversion or ADM
  - SOFA
  - SARB
  - TISA
- Code organized into 12 Subsystems
  - Each subsystem tied to 1 or more working groups
- Each Subsystem made up of 1 or more Product Generation Executives (PGEs)
  - Currently there are about 70 active PGEs

Data from other Instruments used by CERES

- CERES Instrument/ERBElike only subsystems that can process when only CERES data available.
- CERES directly uses the following MODIS data sets:
  - MYD02SS1/MOD02SS1* (19 channel radiance subset of every other pixel every other scanline)
  - MYD03/MOD03* (geolocation)
  - MYD04_L2/MOD04_L2 (5 min 10 km aerosol swath)
  - MYD08_D3/ MOD08_D3 (daily 1 deg aerosol)
    - Critical data sets; must have matched pairs to process.
- Additionally CERES uses, Geostationary satellite data:
  - MET-5, MET-6, MET-7, MET-8, MET-9
  - GOES-8, GOES-9, GOES-10, GOES-11, GOES-12
  - GMS-5, MTSAT-1R

CERES Processing Software

Subsystem Number	Subsystem Name	LOC (to nearest 1K)	Publicly Available Date Products	Product Frequency	Comments
	CERESlib	115K			All Satellites
1	Instrument/Pre-Processor	4K			NPP only
1	Instrument	110K	BDS	1/day	All Satellites
2	ERBElike/ Inversion	33K	ES-8	1/day	All Satellites
3	ERBElike/ TSA	16K	ES-9, ES-4	1/month	All Satellites
12	MOA	10K			Run monthly
4.1 – 4.4	Clouds	231K			All Satellites
4.5 – 4.6	Inversion	26K	SSF	1/hour	All Satellites
5	SARB	51K	CRS	1/hour	All Satellites
6 & 9	TISA-Gridding	31K	FSW, SFC, ISCCP-D2like-Day/Nit	60/month, 36/month, 1/month	All Satellites
11	GGEO	50K	ISCCP-D2like-GEO	1/month	Geostationary
7.2	Synoptic SARB	10K			All Satellites
7.1 & 8 10	TISA-Averaging	164K	SYN, AVG, ZAVG SRBAVG	1/day, 1/month, 1/month 5/month	All Satellites
	TOTAL LOC	851K			

Current CERES Terra/Aqua Processing Approach

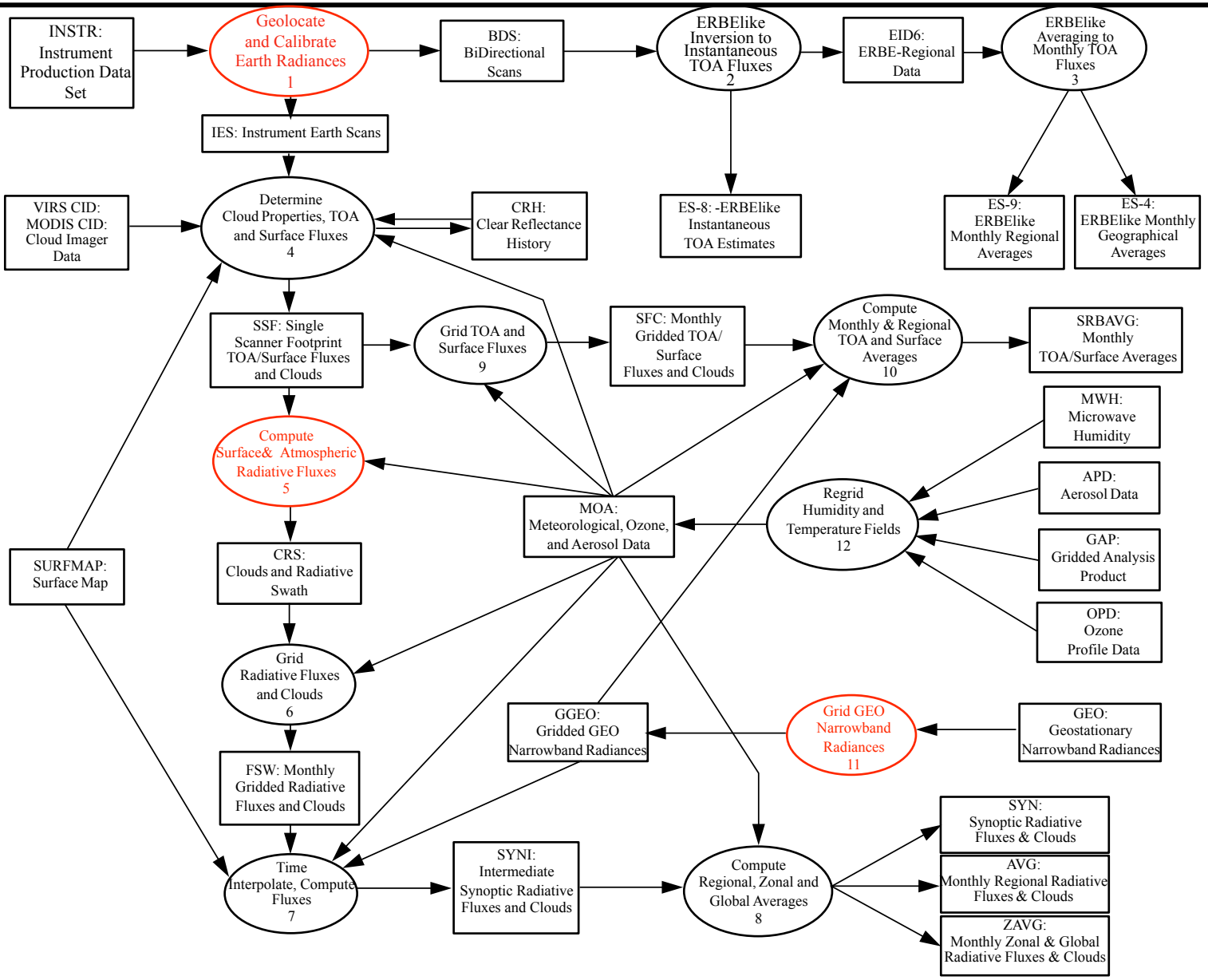
CERES processes data 3 times

Baseline1-QC	**Edition1-CV**	**Edition2**
• Processed daily • Run Instrument & ERBElike Inversion subsystems • Use Composite Snow Map • Not publicly available	• Processed monthly • Run Instrument and ERBElike subsystems • Use actual Snow map and wait for all expected instrument inputs • CV stands for "Calibration/Validation" • Primary Instrument & ERBElike products made publicly available	• Processed in blocks of 4+ months at a time • Run all CERES subsystems as inputs become available • All primary archival products made publicly available

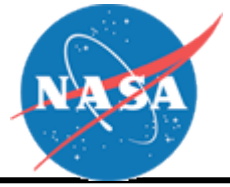
Edition2 Processing

	Stage 1	Stage 2	Stage 3	Stage 4
Latency	6 – 24 months (Wait for Gains and SRF based on Edition1-CV)	~6 months after SFC (Wait for Cal Coef for geo-sat)	Wait for aerosol inputs (MATCH-like data)	Wait for GGEO and FSW availability
Processing	Instrument (BDS, IES)* ERBElike Inv (ES8)* ERBElike TSA (ES4, ES9) MOA (MOA) Clouds (Temp)* Inversion (SSF)* TISA-gridding (SFC)	GGEO (GGEO) TISA-Averaging (SRBAVG)	SARB (CRS)* TISA-Gridding (FSW)	TISA-averaging (TSIB) Synoptic SARB (SYNI) TISA-Averaging (SYN, AVG, & ZAVG) * Instantaneous

CERES Data Flow Diagram



Relevant NPP Issues



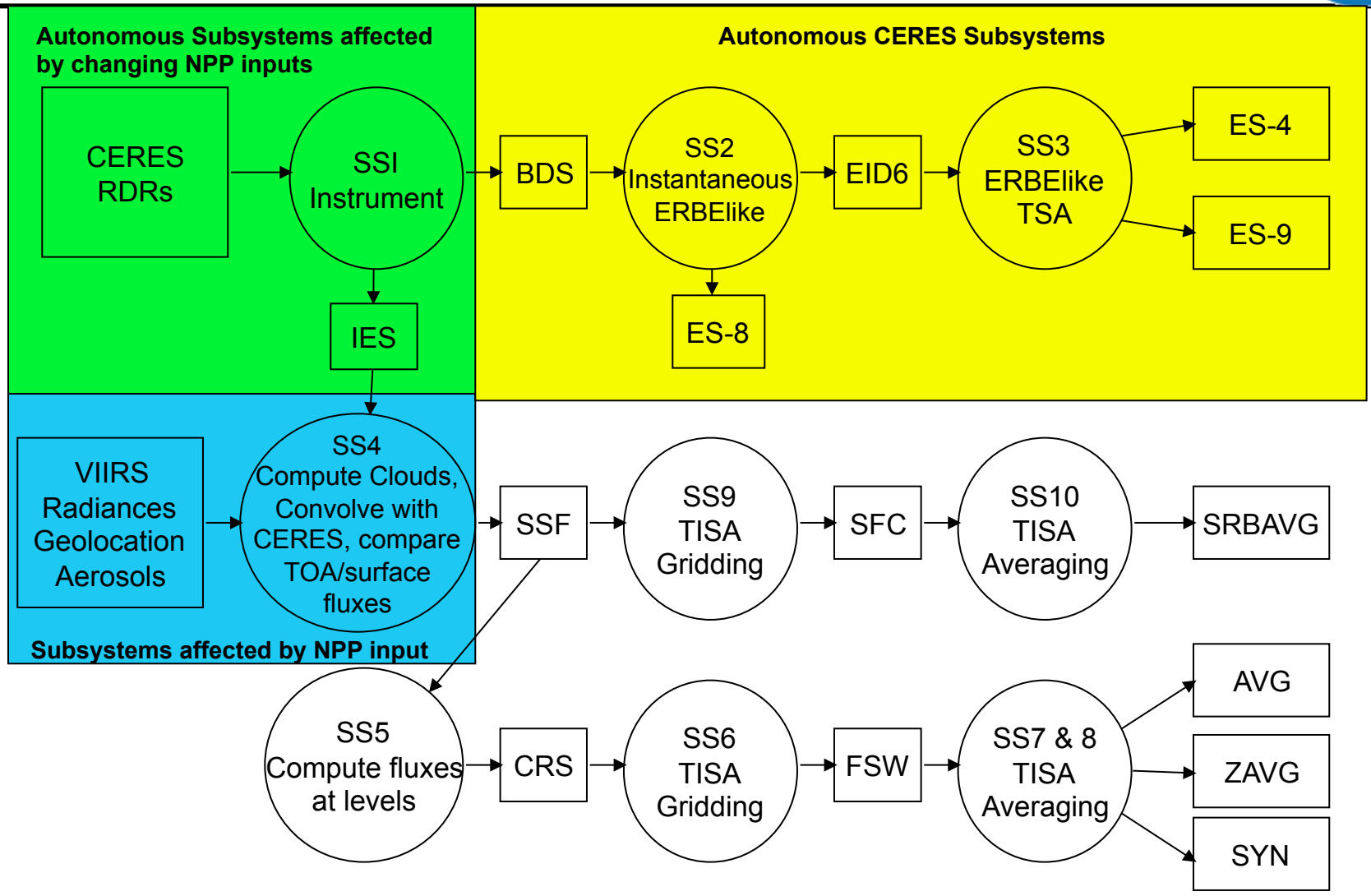
- Imager input data required for Climate Data Record (CDR) must be of climate quality and consistently calibrated over entire period.
  - In NPP era, Land PEATE provides CERES aggregated radiance and geolocation files and sub-sampled data files using CERES provided code. Land PEATE also provides AOT files that correspond to sub-sampled radiance/geolocation.
  - For Terra/Aqua, MODAPS provides radiance, geolocation, and aerosol files from a collection that begins at covers open.
- NPP CERES made use of already existing interfaces.
  - Cost savings by using existing infrastructure.
  - Land PEATE already getting VIIRS data. Agreed to also obtain CERES RDRs.
  - Network between Land PEATE and ASDC exists for Terra/Aqua.

CERES Input Data

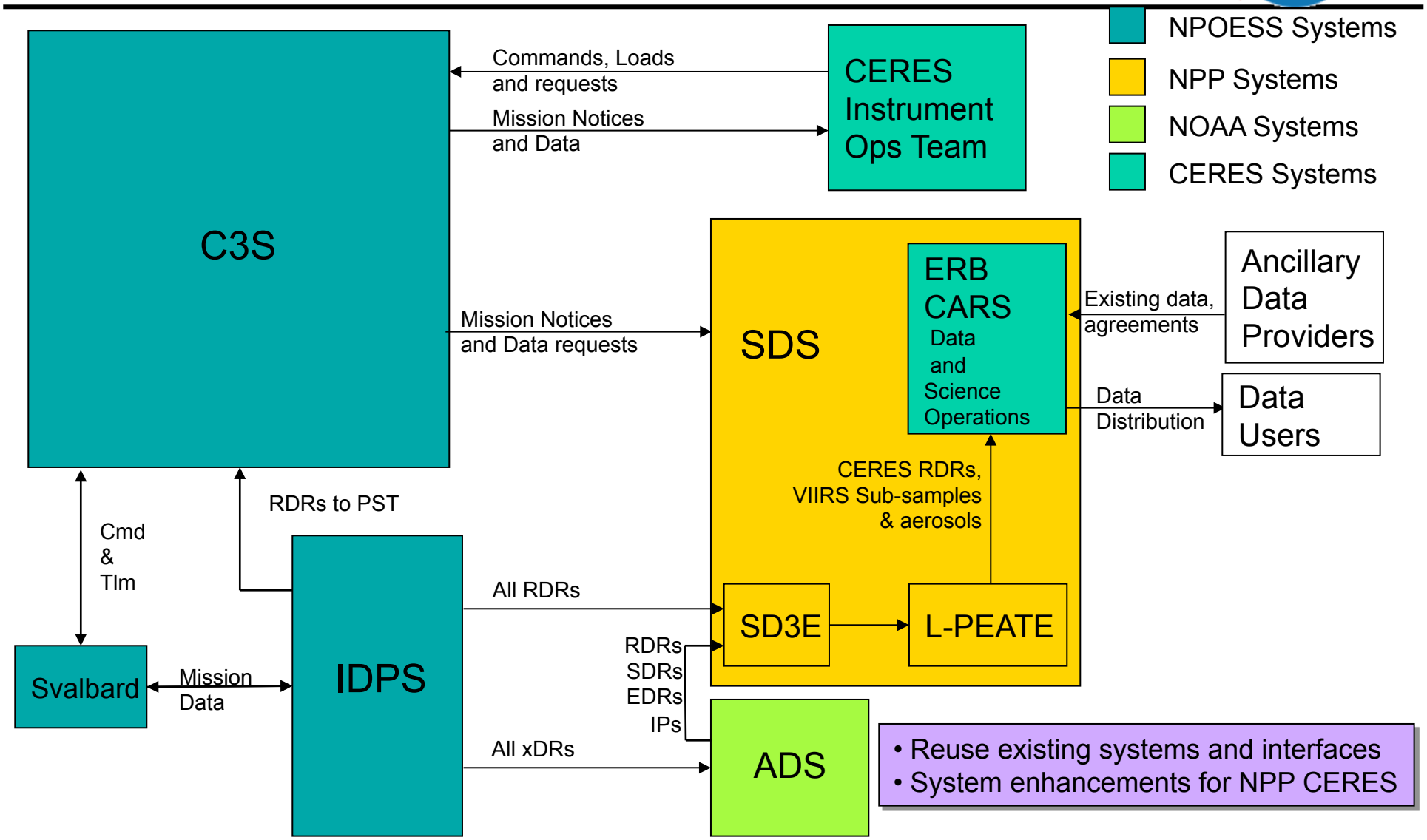
Type of Data	Parameter Description	Terra/Aqua Freq & Source	NPP Freq & Source	Comments
CERES L0 files	Instrument level 0 data	3/day; EDOS @ GSFC	~131/day; Land PEATE	In case of NPP, RDRs also contains spacecraft diary
Attitude	Attitude	12/day; GSFC Flight Dyn Facility	included in RDR	
Ephemeris	Ephemeris	12/day (Terra); GSFC Flight Dyn Facility 1/day (Aqua); same	included in RDR	
Imager Calibrated Data, Instantaneous	Imager Radiances & Geolocation Aerosols	288/day; MODAPS ~144/day; MODAPS	288/day; Land PEATE ~144/day; Land PEATE	CERES provided code to sub-sample radiance files at GSFC
Aerosol data	Aerosol (Coln) Optical thickness, type/size	1/day; MODAPS		For Terra/Aqua using MODIS MOD08 and MATCH. Plan to do same for NPP
Meteorological and Ozone data	3-D Met Data 2-D atmospheric data 2-D constants	4/day; GMAO 24/day; GMAO 1; GMAO		
Precipitable Water	2-D constants	2/day; Global Hydrology Resource Center (GHRC)		
Geostationary data	MCIDAS data from 5 geostationary satellites per month	120/day; University of Wisconsin Space Science and Engineering Center (SSEC)		Only every 3rd hour is used for production
SURFMAP(Snow/Ice)	Snow/Ice Map	4/day; NCEP/NESDIS		
SURFMAP(Snow/Ice)	Snow/Ice Map	1/day; NSIDC		

Shared NPP and Terra/Aqua Data Source

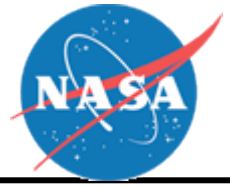
Simplified CERES Processing Flow



NPP CERES Operational Data Flow



FM5 Code Development

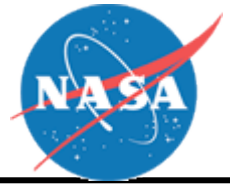


- Instrument only subsystem modified
  - Preprocessor will convert data to format consistent with Terra and Aqua
  - Instrument subsystem current implemented with Ada
  - Convert Ada to C++ and deliver preprocessor for Ada as schedule risk mitigation
- Six total code deliveries
  - Deliver Ada FM5 ready December 4 (AMI x86)
  - Deliver C++ FM5 ready March 12 (AMI P6)

Main Terra and Aqua Edition2 Data Sets

Product	Latest Edition	Data available through	Comments
BDS, ES8, ES4, ES9	Edition2 (T, A)	Dec'08	Waiting on gains/SRF
SSF	Edition2F (T) Edition2C (A)	Dec'07	Processing Terra through 2008
SFC	Edition2F (T) Edition2C (A)	Dec'07	
SRBAVG	Edition2D (T) Edition2A (A)	Oct'05	Waiting on MTSAT coefficients
CRS	Edition2F (T) Edition2C (A)	Dec'07	Have MATCH inputs thru Dec'07.
FSW	Edition2F (T) Edition2C (A)	Dec'07	
SYN, AVG, ZAVG	Edition2C (T) Edition2B (A)	Oct'05	Waiting on MTSAT coefficients
ISCCP-D2like-Day, ISCCP-D2like-Nit	Beta1	Aug'07	Awaiting Edition 2 Code delivery this month
ISCCP-D2like-Geo	Beta1	Oct'05	Awaiting Edition 2 code delivery this month

Terra and Aqua Edition 3

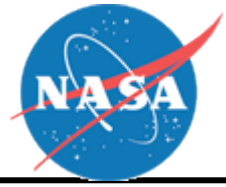


- Terra Edition 3 instrument ready to process this week
- Aqua Edition 3 data sets are delayed until Gains and Spectral Response Functions (SRF) available
  - Once Science approves Gains & SRF, production expected to begin within 4-6 weeks
  - Other CERES Edition3 data sets will follow as code deliveries are made
- CERES will reprocess and forward process all data
- Product parameters may be added or changed
  - SSF size to increase over 33%! (over 50 parameters added)
- Edition2 will extend until Edition3 catches up
  - Expect letter change in Edition2 data set names starting 2008 due to switch from GEOS-4 to G5-CERES

Terra and Aqua Edition 3

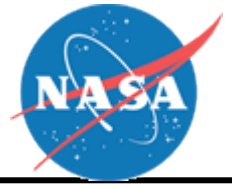
Subsystem	Data Start	Data End	Processing Complete
(T) Instrument	Mar'00	Dec'04	1/30/2010
(T) ERBE-like	Mar'00	Dec'04	3/5/2010
(A) Instrument	June'02	Mar'09	5/21/2010
(A) ERBE-like	June'02	Mar'09	5/28/2010
(T) Instrument	Jan'05	Mar'09	3/26/2010
(T) ERBE-like	Jan'05	Mar'09	3/26/2010

Development and Production Platforms



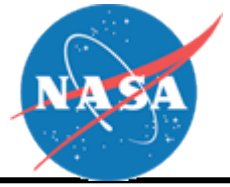
- Most efficient to develop code on machine that is identical to production platform
  - Same environment including Toolkit, operating system, compiler, HDF, and library versions
  - Science approves for delivery to production platform newly developed code by examining associated data runs
    - Science shouldn't have to repeat exercise on production platform
    - Unless production environment itself introduced a change, data management can quickly compare files created in development and production environments
- Access to production output products from development machine improves efficiency
  - No need to order data, no duplicate copies of products
  - Faster to evaluate and use data sets

Lessons Learned When Delivering Code and Testing



- Data Management personnel create expected output on target production machine
  - Verify that expected output looks as expected on target machine
- CM untars delivery and compiles source code on target machine, runs to reproduce expected output
  - Verify that delivery tar file includes all necessary components prior to turning delivery over to ASDC for testing
- Once delivery is in production do not immediately begin running an Edition data set
  - Run ValRx for all instrument/input combinations
    - Ensure production environment not altering output
    - Ensure correct files were delivered
    - Ensure scripts set up correctly

Platform Migration

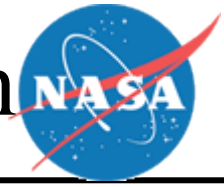


- Codes typically must be modified to work on new platform
  - Extensive updates may be needed
  - Takes time, may not be highest priority
  - Currently migrating last 4 CERES subsystems off SGI and onto IBM cluster
- Because CERES produces Climate Data Records, must verify that output is scientifically equivalent regardless of production platform
  - Can't upgrade algorithms as part of migration

AMI Transition

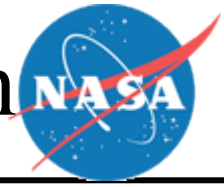
Subsystem	Edition	Target Delivery to AMI
Instrument	Ed2	12/4/2009
ERBE-like	Ed2 & Ed3	6/25/2010
Clouds Convolution	Ed3	6/11/2010
Inversion SOFA	Ed2	8/20/2010
Instantaneous SARB	Ed2	3/5/2010
Synoptic SARB	Ed2	5/14/2010
TISA GGEO	Ed3	4/9/2010
TISA SRBAVG, AVG, ZAVG, SYN	Ed2	4/30/2010
TISA SFC & FSW	Ed2	3/19/2010
MOA	Ed3	12/14/09
PMOA	Ed3	2/19/2010

Production Processing Optimization



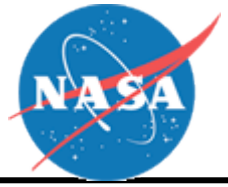
- Increased capacity with AMI environment: Utilize with optimized code and production scheduling
- Code optimization
  - Plan to work with ODU beginning in November to explore parallelization techniques to implement within source code.
  - TISA software is first test case (AVG, ZAVG, SYN)
  - Working Groups will decide maintainability of suggested source code changes

Production Processing Optimization



- Production requests, job submittal, production monitoring and reporting currently independent
- Migrate PRs to relational database
  - Dependencies for PGE represented
- Develop tools to retrieve and display production status
  - Currently manually search archive for output
- Ideally store production status in database for PR generation
- Maximize scheduling with Sun Grid Engine

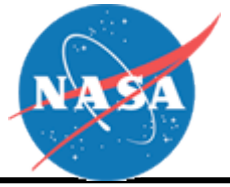
CERES uses CMMI Approved Processes



- The CERES DM task was successfully appraised at CMMI Maturity Level 2. All individual Process Areas were appraised at Capability Level 3.
- CERES DM task CMMI processes are described in 9 process plans (http://science.larc.nasa.gov/ceres/DMP_Plans/index.html)
  - Data Management Process Plans - Requirements Management Plan
  - Configuration Management Plan - Risk Management Plan
  - Data Management Plan – DRAFT - Software Development Plan
  - Measurement and Analysis Plan - Software Management Plan
  - Process and Product Quality Assurance Plan - Training Management Plan
- During the recent (Aug 12th) ISO9100 internal process audit of the (entire) Science Directorate, CERES was audited as a representative project. There were NO non-conformances identified.

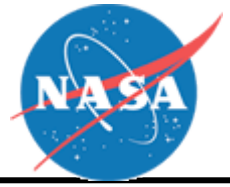
Documentation Overview

(http://eosweb.larc.nasa.gov/PRODOCS/ceres/table_ceres.html)



- Data Quality Summaries
  - Detailed information about a particular data set
  - Living Document; most up-to-date
  - Always consult Data Quality Summary prior to using data or publishing research
- Data Products Catalog
  - Parameter lists for each data product
  - Version of pages that apply to data set included with order
- Collection Guides
  - User Guide for data product
- Description/Abstract
  - Record of differences between data sets and configuration codes

Questions about Data Sets??

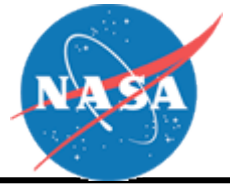


- Look over Data Products Catalog pages
- Reread Data Quality Summary
- Consult Collection Guide, if available
- Specific science questions may be sent to Contact Scientist listed in Section 2.2 of Collection Guide or in Description/Abstract
- All other questions should be sent to User Services

  larc@eos.nasa.gov
- For data products for which no Collection Guide or Description/Abstract is available, send all questions to User Services

Science Data Product URLs and Contacts

- Ordering Data
  - http://eosweb.larc.nasa.gov/HBDOCS/langley_web_tool.html
  - https://wist.echo.nasa.gov/api
  - EOS Data Gateway was decommissioned Feb. 27, 2009.
- Subsets of SSF, CRS, and ES8 are available
  - Order data using Java version of Langley Ordering Tool
  - Can subset by parameters or latitude/longitude box
- Contact Points
  - All questions regarding production data products and their use
    - E-mail: larc@eos.nasa.gov
    - Langley ASDC Customer Service
- CERES News (e-mail)
  - Subscribe from CERES Data Products webpage
  - All new public datasets are announced soon after public release
  - Mechanism for distributing CERES information



Questions and Comments