

CERES Data Management Status

Presented to
CERES Science Team

Mike Little
Erika Geier

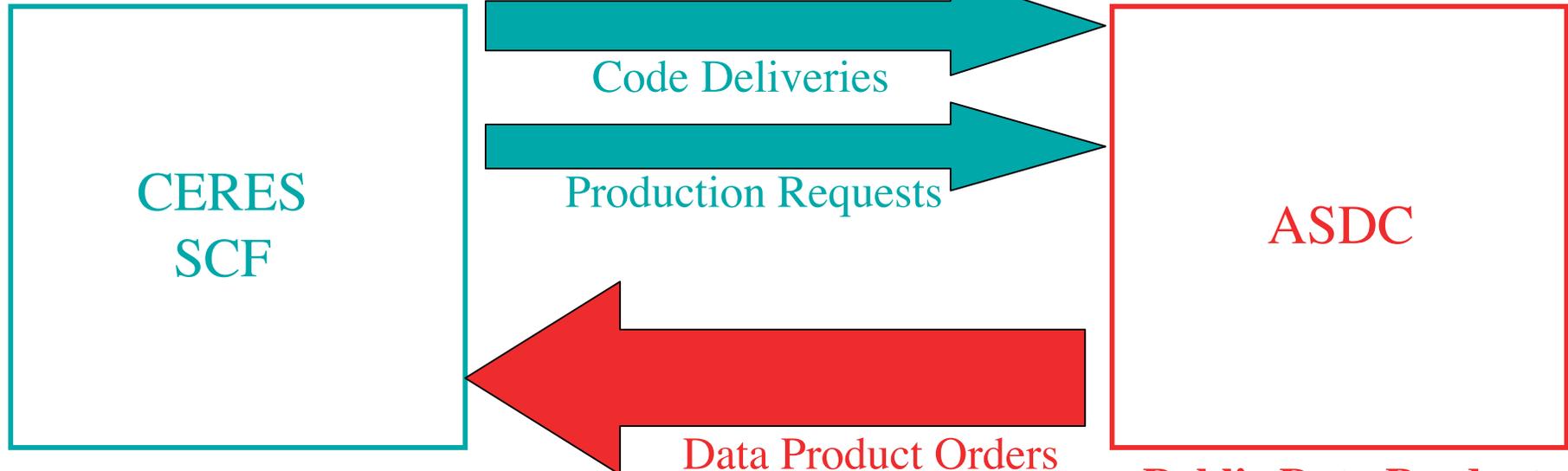
NASA Langley Research Center
May 2, 2006

michael.m.little@nasa.gov

erika.b.geier@nasa.gov

CERES DM Relationship to ASDC

- Configuration Managed Code
- Documentation
- Sample Read Packages



CERES
SCF

Code Deliveries

Production Requests

ASDC

Data Product Orders

Support Science Working Groups

- Algorithm Development
- Code Development
- Calibration/Val analysis
- Data Analysis
- Special Studies

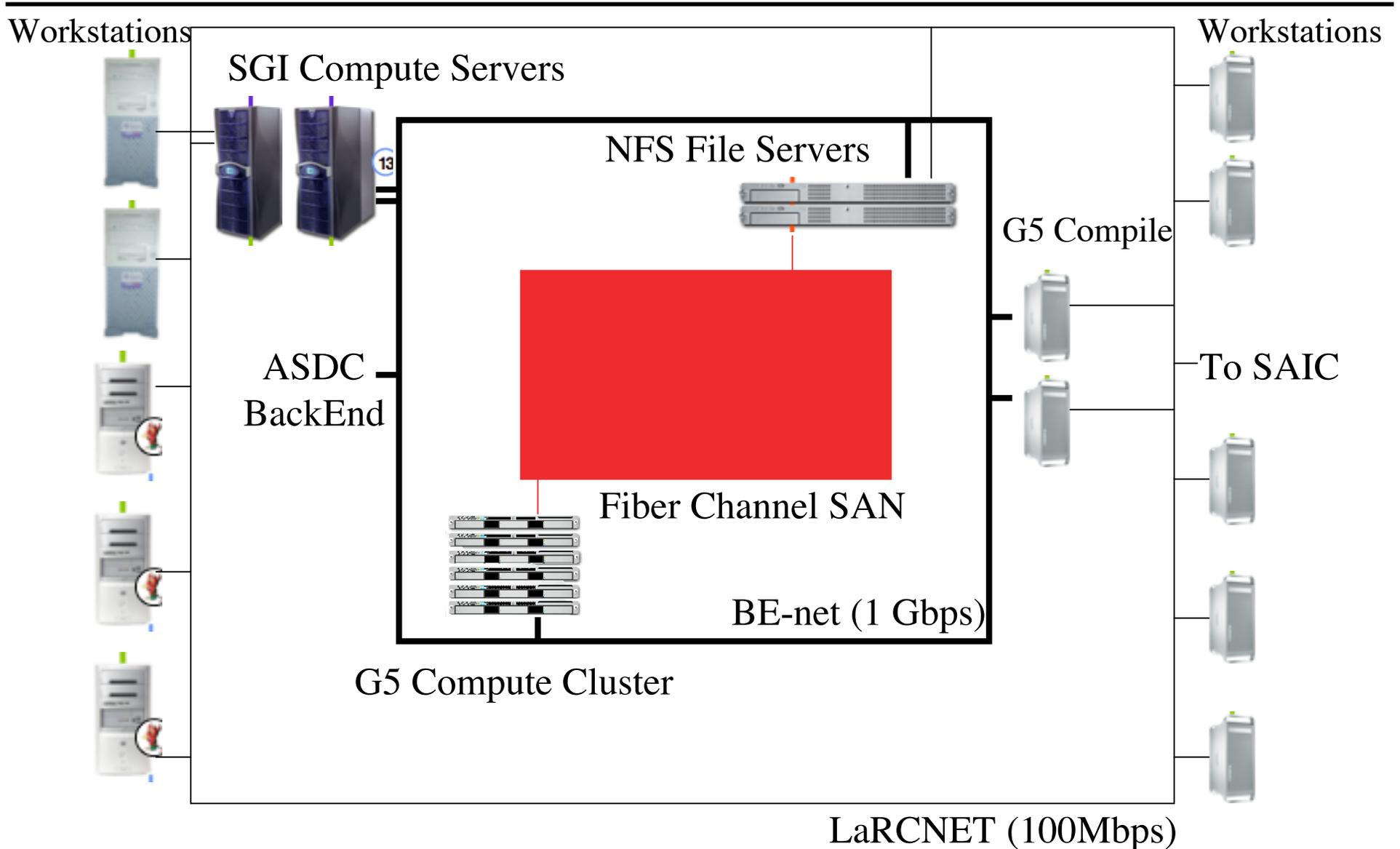
Public Data Products

- QA Data during Production
- Status of Production Requests
- Data Product Orders
- Access to data on Disk Array
- Production
- Reprocessing
- Archive
- Distribution

CERES Data Management is Evolving

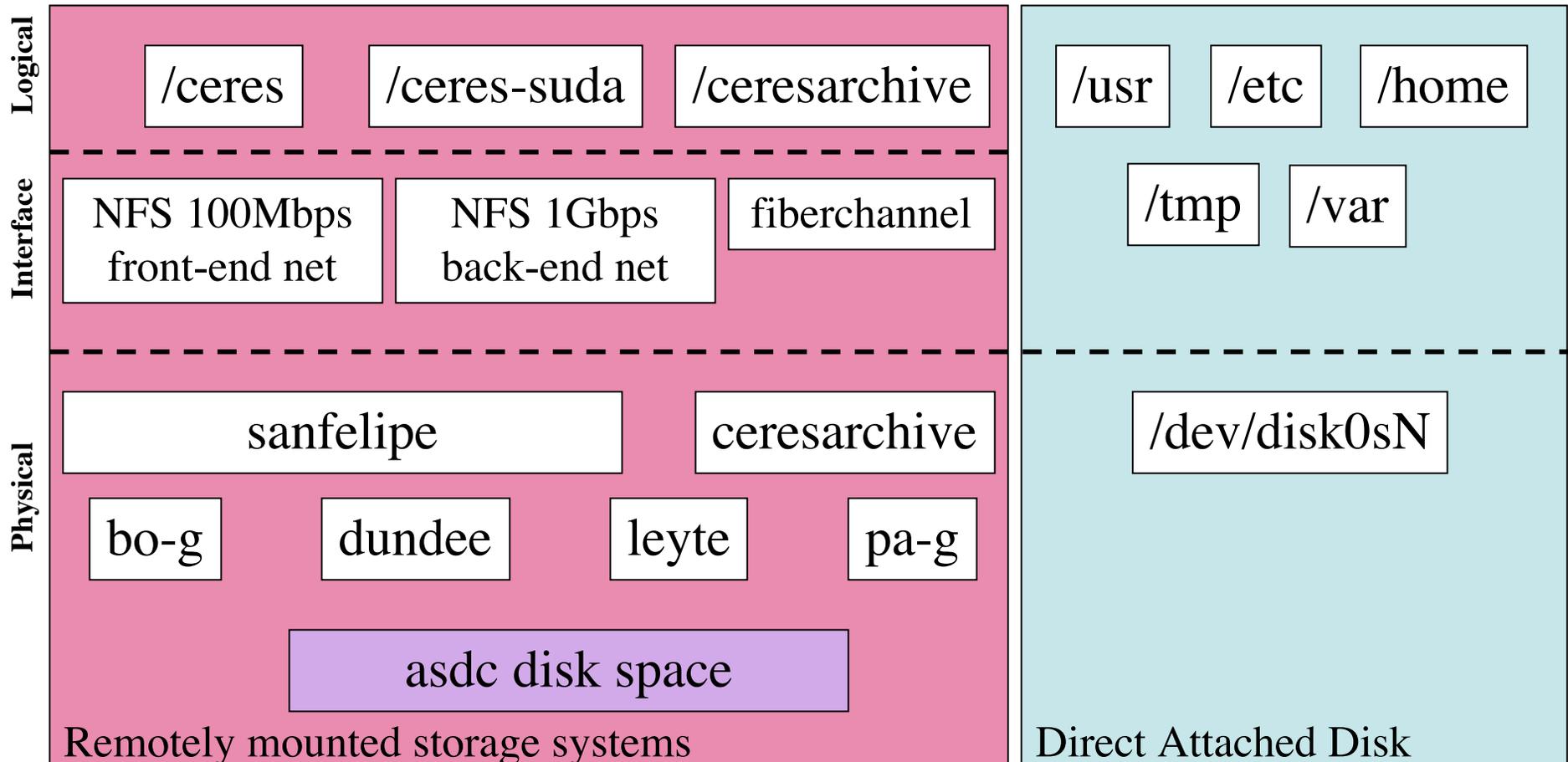
- Need to change driven by
 - ASDC Evolution
 - Cost of IT support to CERES Science/Validation
 - Improved coordination with ASDC
 - Need to support smaller R&A CERES activities
- It's ALL about the Data
 - Upgrade storage for 20-30 instrument-years of data
 - Replace computer-centric with data-centric view of environment
- Changes Needed in Data Management and SCF
 - Storage Area Network makes same data available everywhere
 - Convert Codes/Scripts to run on multiple platform-types
 - Modernize SCF to use commodity cluster/grid computing
 - Computing Hardware
 - Scheduling Tools
 - Cultural Changes

Target CERES SCF Architecture



Logical File Architecture

/



Physical Storage Architecture

	sanfelipe		ceresarchive	ceres-temp	Direct Attached Disk
Purpose	ASDC data online	SciTeam only shared data	Non-public long term retention	Cheap, quick reaction to need	Computation output Scratch or interactive
Latency	Low to cluster		Moderate	Moderate	Low
Reliability	High		High	Moderate	Moderate
Back-up	At ASDC		Yes	Yes	Yes
Read/Write	•Tsar-Write •Users-Read	•Tsar allocate •Users-R/W	•Tsar allocate •Users - R/W	•Tsar allocate •Users - R/W	•User defined & populated
Sharing	•FC-zamboanga •FE-Network, BE-Network		•FC-zamboanga •BE-Network	•FE Network •BE-Network	•Not generally •FE-Network
Metadata	•At ASDC	•Yes	•Yes	•No	•No
Physical	•xServe RAID, xSAN		•xServe RAID,xSAN	•LaCie, G4	•Internal disk drive
Capacity-06	•40TB	•40TB	•10 TB	•10TB	•80-100GB-Workstation •100GB-Cluster node •10TB-thunder/lightning

CERES Code Conversion

- Code Conversion to run on Multiple Platforms
 - Conducted Initial Experiments to define Process
 - Debug computational environment
 - Resolve Compiler Issues
 - Document Process and Lessons Learned
 - Define Requirements for Code Certification for a Platform
 - What is Scientific Identity for each Subsystem?
 - Invariance across platforms, OS, processors
 - Maximize automation of certification for all changes across all platforms
 - Configuration Management of the Codes at the version level
 - Both PGEs and scripts
 - Needed to preserve certification
- Cluster Management Tools
 - Resource Manager - Sun Grid Engine
 - Job Scheduling - 2 Modes
 - On-demand processing
 - “Run like it was on the ASDC” - S4PM
 - Close coordination with ASDC Evolution Program
 - Requirements for Changes to ASDC Uncertain

CERES SCF Cluster

- Storage Services
 - Current data is mostly identified with a specific computer
 - Workstation is user's personal storage
 - Storage Area Network provides high volume, high reliability, shared storage
 - Common data accessible from all platforms in SCF
 - Decouple physical location from logical directory structure
 - Permit codes to access data from any machine without change
 - Backup and Archive all "interesting" data
 - Some limited scratch space
 - Minimize duplication of files
- Compute Services
 - SGI retirements start about May 07 with full phase-out by May 09
 - Mac G5 cluster to provide a "compute appliance"
 - Add capacity to replace SGI capacity
 - Expand to support new projects as they come along
 - Several Mac G5 servers to provide several interactive machines
 - Compilation, small jobs, testing, etc
 - Workstations currently cannot run most codes or reliably compile source
 - Future configurations to permit compilation and testing compatible with clusters

CERES Cultural Changes

- **Code Management**
 - Configuration Management at the version level to ensure audit trail of certification
 - Coordinate with ASDC delivery process
- **Retention of Data on Storage system**
 - Storage Area Network provides high volume, high reliability, shared storage
 - Eliminate duplicate files, particularly Data Products available at ASDC
 - Maintain metadata for all files/directories for future review/elimination
- **Delivery Process to Support CERES Automation at ASDC**
 - See ASDC presentation
- **Re-allocate jobs between ASDC and SCF**
 - ASDC to focus on data product production
 - Reduce pre-production testing and special runs
 - SCF to run ValRx, short runs, etc.
 - Need SCF to be capable of “Running Jobs Like the ASDC Does”

CERES Evolution Status

- **Code Conversion (SGI, Mac, IBM)**
 - Cereslib, SARB, Instrument, Synoptic SARB, Regrid MOA done
 - ERBE-like PGEs done, scripts in progress
 - Clouds started
 - Convolution converted
 - Other subsystems will convert not to interfere with Algorithm work
 - Expect to be finished by Summer, 2007
- **Storage System**
 - Physical reconfiguration is in progress during this meeting
 - Need to establish a storage Tsar to monitor/allocate resources
 - Need to split clouds/convolution group into 3 parts
 - Capacities
 - SGI-10TB
 - SAN-19TB (35 TB to be added this month)
 - Cheap Seats - 14 TB
- **Communications System**
 - Improvements to Backend network (1Gbps file services) expected in Jun, 2006
 - Improvements to Fiberchannel network waiting on hardware delivery
 - Front-end network (100Mbps) is in place

CERES Evolution Status

- Compute Services
 - Hardware
 - 7-node cluster in place and tested
 - Initial Compilation servers in place and tested
 - Some workstations in place for testing
 - 8 nodes for CALIPSO, A-Train delivered yesterday
 - 30 additional nodes on order to replace thunder and lightning
 - Resource Manager (SGE) set up
 - Different queues for different Projects, priorities
 - Job Scheduler
 - Adhoc job scheduler is available
 - “Run like the ASDC” S4PM job scheduler being coordinated with ASDC
- Agreement with ASDC for CERES Automation
 - Agreement in principle
 - Details to be worked out
 - More description during ASDC talk tomorrow

The Big News: Edition3 is coming!!

- Edition2 expected to end with December 2005 data
- Edition3 expected to start processing in Summer 2006
 - Phase-in slowly, starting with BDS and ERBElike data products
 - Parameters on product may be added or changed
 - Terra and Aqua data sets will be of very similar quality
 - MODIS collection 5 input to Terra/Aqua cloud processing
 - Reprocess all data and do forward processing
 - Beta data set to run first for most products

CERES Science Data Sets

New Since November 2005 (SSF, SFC, and SRBAVG)

Data Set	Comments
CER_SRBAVG_Terra-FMx-MODIS_Edition2D	<ul style="list-style-type: none">• Edition2C SFC are input• 3 years processed (3/00 – 2/03)• To extend through 5/04 this summer• Available to Science Team; imminent public release
CER_SRBAVG_Terra-FMx-MODIS_Beta6	<ul style="list-style-type: none">• Edition2C SFC are input• 3/03 - 5/04 processed• Available to Science Team only (no public release)
CER_SSF_Aqua-FMx-MODIS_Edition2A	<ul style="list-style-type: none">• Aqua ADMs, updated Surface LW Model B Flux• 7/02 - 3/05 processed• To extend through 12/05 this summer• Publicly available
CER_SFC_Aqua-FMx-MODIS_Edition2A	<ul style="list-style-type: none">• Edition2A SSF are input• 7/02 – 2/05 processed• To extend through 12/05 this summer• Available to Science Team; public release expected

CERES Science Data Sets

New Since November 2005 (Instrument and ERBElike)

Data Set	Comments
CER_BDS_Terra-FMx_Edition1-CV CER_ES8_Terra-FMx_Edition1-CV CER_ES4_Terra-FMx_Edition1-CV CER_ES9_Terra-FMx_Edition1-CV CER_BDS_Aqua-FMx_Edition1-CV CER_ES8_Aqua-FMx_Edition1-CV CER_ES4_Aqua-FMx_Edition1-CV CER_ES9_Aqua-FMx_Edition1-CV	<ul style="list-style-type: none"> • All Terra and Aqua FM3 data from time the covers opened through present • Aqua FM4 data from 6/19/02 - 3/29/05 • Calibration/validation data sets that replace Edition1 • Algorithms to be held constant across entire data set • Publicly release
CER_BDS_Aqua-FM4_Ed1-CV-NoSW CER_ES8_Aqua-FM4_Ed1-CV-NoSW CER_ES4_Aqua-FM4_Ed1-CV-NoSW CER_ES9_Aqua-FM4_Ed1-CV-NoSW	<ul style="list-style-type: none"> • 3/30/05 through present • Same algorithms and processing as Edition1-CV • Due to SW channel failure, no daytime SW or LW radiances and fluxes • Available to Science Team; public release expected
CER_BDS_Aqua-FM4_Ed2-NoSW CER_ES8_Aqua-FM4_Ed2-NoSW CER_ES4_Aqua-FM4_Ed2-NoSW CER_ES9_Aqua-FM4_Ed2-NoSW	<ul style="list-style-type: none"> • 3/30/05 through 1/1/06 • Same algorithms and processing as Edition2 • Due to SW channel failure, no daytime SW or LW radiances and fluxes • Available to Science Team; public release expected
CER_BDS_Aqua-FM4_AnomOps-Ed1-CV CER_ES8_Aqua-FM4_AnomOps-Ed1-CV CER_ES4_Aqua-FM4_AnomOps-Ed1-CV CER_ES9_Aqua-FM4_AnomOpsEd1-CV	<ul style="list-style-type: none"> • 7/19/05 through 8/30/05 • Same algorithms and processing as Edition2 • Due to instrument temperature changes made in hopes of reversing SW channel problem, data quality varies from Ed1-CV-NoSW • Available to Science Team only (no public release)

CERES Science Data Sets

Expected Before Next Science Team Meeting (Editions 1 & 2)

Data Set	Comments	When?
CER_SSF_Aqua-FM4-MODIS_Ed2A-NoSW CER_SFC_Aqua-FM4-MODIS_Ed2A-NoSW	<ul style="list-style-type: none"> • 3/30/05 hour 18 through 1/1/06 • Same algorithms and processing as Edition2 • Due to SW channel failure, no daytime SW or LW • SSF publicly available when processed; SFC public release expected 	To start May 2006
CER_SRBAVG_Aqua-FMx-MODIS_Beta5	<ul style="list-style-type: none"> • Edition2A SFC are input • Process 7/02 – 2/03 • Available to Science Team only (no public release) 	To start shortly
CER_CRS_Aqua-FMx-MODIS_Edition2A	<ul style="list-style-type: none"> • Aqua Edition2A SSF are input • Process all month for which inputs available (initially 7/02 - 3/05) • Public release expected 	To start as soon as SARB approves validation runs
CER_FSW_Aqua-FMx-MODIS_Edition2A	<ul style="list-style-type: none"> • Aqua Edition2A CRS are input • Process all month for which inputs available • Public release expected 	To start Summer 2006
CER_BDS_TRMM-PFM_Edition1-CV CER_ES8_TRMM-PFM_Edition1-CV CER_ES4_TRMM-PFM_Edition1-CV CER_ES9_TRMM-PFM_Edition1-CV CER_BDS_TRMM-PFM_Ed1-CV-TransOps CER_ES8_TRMM-PFM_Ed1-CV-TransOps CER_ES4_TRMM-PFM_Ed1-CV-TransOps CER_ES9_TRMM-PFM_Ed1-CV-TransOps	<ul style="list-style-type: none"> • 12/27/1997 through 8/31/1998; March 2000 (Edition1-CV) • Select transient operation periods during 1999 (Ed1-CV-TransOps) • Calibration/validation data sets that replace Edition1 • Algorithms to be held constant across entire data set • Algorithms consistent between Terra and Aqua • Public release expected 	To start processing as resources become available (very low priority)

CERES Science Data Sets

Expected Before Next Science Team Meeting (Edition3)

Data Set	Comments	When?
CER_BDS_Aqua-FMx-MODIS_Edition3 CER_BDS_Terra-FMx-MODIS_Edition3 CER_ES8_Aqua-FMx_Edition3 CER_ES8_Terra-FMx_Edition3 CER_ES4_Aqua-FMx_Edition3 CER_ES4_Terra-FMx_Edition3 CER_ES9_Aqua-FMx_Edition3 CER_ES9_Terra-FMx_Edition3	<ul style="list-style-type: none"> • The first Edition3 data sets! • Terra to process 2/25/00 forward • Aqua to process 6/18/02 forward 	To start Aug/Sept 2006
CER_SSF_Terra-FMx-MODIS_BetaEd3 CER_SFC_Terra-FMx-MODIS_BetaEd3 CER_SSF_Aqua-FMx-MODIS_BetaEd3 CER_SFC_Aqua-FMx-MODIS_BetaEd3	<ul style="list-style-type: none"> • Beta run of proposed Edition3 data sets • 6 to 8 months of data expected to be processed for each instrument • Many new parameters • Some changed and deleted parameters 	To start Aug/Sept 2006

Older CERES Science Data Sets

Still of Interest (TRMM)

Data Set	Comments
CER_BDS_TRMM-PFM_Edition1 CER_ES8_TRMM-PFM_Edition2 CER_ES4_TRMM-PFM_Edition2 CER_ES9_TRMM-PFM_Edition2 CER_SSF_TRMM-PFM-VIRS_Edition2B CER_SFC_TRMM-PFM-VIRS_Edition2B CER_SRBAVG_TRMM-PFM-VIRS_Edition2B CER_CRS_TRMM-PFM-VIRS_Edition2C CER_FSW_TRMM-PFM-VIRS_Edition2C	<ul style="list-style-type: none"> • TRMM period Jan'98 – Aug'98; Mar'00 • Publicly available
CER_BDS_TRMM-PFM_Transient-Ops2 CER_ES8_TRMM-PFM_Transient-Ops2 CER_SSF_TRMM-PFM-VIRS_Edition2B-TransOps	<ul style="list-style-type: none"> • Transient Operations period Jan'99 – Jul'99 • BDS not publicly released; others publicly available
CER_ES4_PFM+FM1+FM2_Edition2 CER_ES9_PFM+FM1+FM2_Edition2 CER_ES4_PFM+FM1_Edition2 CER_ES9_PFM+FM1_Edition2 CER_ES4_PFM+FM2_Edition2 CER_ES9_PFM+FM2_Edition2	<ul style="list-style-type: none"> • Mar'00 only month of TRMM, Terra overlap • Publicly available
CER_SSF_TRMM-SIM-VIRS_Edition2-VIROnly	<ul style="list-style-type: none"> • Processed only for months when CERES was off • Processed through Jul'01; no additional months expected • Publicly available

Older CERES Science Data Sets

Still of Interest (Terra)

Data Set	Comments
CER_BDS_Terra-FMx_Edition2 CER_ES8_Terra-FMx_Edition2 CER_ES4_Terra-FMx_Edition2 CER_ES9_Terra-FMx_Edition2	<ul style="list-style-type: none"> • Rev1 SW corrections available • Terra Edition2 BDS and ES8 processed 2/25/00 – 6/30/05 • Terra Edition2 ES4 and ES9 processed 3/00 – 5/05 • Data set to be extended through 12/05 shortly • Publicly available • Edition1 (10/05 is last processed month) publicly available
CER_ES4_FM1+FM2_Edition2 CER_ES9_FM1+FM2_Edition2	<ul style="list-style-type: none"> • Processed Mar'00 through Dec'02 • Production of these data sets halted • Replaced by Terra/Aqua combined Crosstrack data sets (FM1+FM3, FM1+FM4 Edition2) • Publicly available
CER_SSF_Terra-FMx_Edition2B CER_SFC_Terra-FMx-MODIS_Edition2C	<ul style="list-style-type: none"> • Rev1 SW corrections available • SSF: 2/25/00 – 6/30/05 processed; 7/1/04 - 1/1/06 still expected • SFC: 3/00 - 5/05 processed; 6/05 - 12/05 still expected • Uses Edition2 instrument data as input • Edition2 Clouds; Final Terra ADMs • Publicly available
CER_CRS_Terra-FMx_Edition2B CER_FSW_Terra-FMx-MODIS_Edition2C	<ul style="list-style-type: none"> • CRS: 3/1/00 – 6/30/05 Crosstrack processed; 7/05 - 12/05 expected • FSW: 3/00 - 5/05 crosstrack processed; 6/05 - 12/05 expected • Edition2B SSF input; latest version of SARB algorithms • Publicly available

Older CERES Science Data Sets

Still of Interest (Aqua)

Data Set	Comments
CER_BDS_Aqua-FMx_Edition2 CER_ES8_Aqua-FMx_Edition2 CER_ES4_Aqua-FMx_Edition2 CER_ES9_Aqua-FMx_Edition2	<ul style="list-style-type: none"> • Rev1 SW corrections available • Aqua-FM4 Edition2 processed 6/18/02 – 3/29/05; Ed2-NoSW begins 3/30/05 • Aqua-FM3 Edition2 processed 6/18/02 – 12/1/05; 12/05 to be processed shortly • Edition2 publicly available; Ed2-NoSW only available to Science Team at this time • FM3 Edition1 (10/05 is last processed month) publicly available • FM4 Edition1 through 3/29/05 publicly available and Ed1-NoSW for 3/30/05 through 11/05 available to Science Team
CER_ES4_FM1+FM3_Edition2 CER_ES4_FM1+FM4_Edition2 CER_ES9_FM1+FM3_Edition2 CER_ES9_FM1+FM4_Edition2	<ul style="list-style-type: none"> • Combine only Crosstrack instruments • Jul'02 – Feb'05 currently available • Uses Terra Edition2 and Aqua Edition2 inputs • Publicly available
CER_SSF_Aqua-FMx-MODIS_Edition1B CER_SFC_Aqua-FMx-MODIS_Edition1B	<ul style="list-style-type: none"> • Rev1 SW corrections available • 6/18/02 – 3/31/05 processed; no additional months will be processed • Uses Edition2 instrument data as input • Terra ADMs used • Publicly available
CER_CRS_Aqua-FMx-MODIS_Beta1 CER_FSW_Aqua-FMx-MODIS_Beta1	<ul style="list-style-type: none"> • First SARB data sets for Aqua • Edition1B SSF (Crosstrack instrument only) is input • 7/02 – 12/04 processed; no additional months will be processed • CRS publicly available; FSW available to Science Team

Documentation Overview

- Data Quality Summaries
 - Detailed information about a particular data set
 - Always consult 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 Set

- 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
 - Can also order data from EOS Data Gateway
- 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 datasets webpage
 - All new public datasets are announced as soon as available
 - Mechanism for distributing CERES information