CERES-II Data Management Status

Presented to
CERES-II Science Team

Mike Little
Erika Geier
NASA Langley Research Center
March 29, 2004

michael.m.little@nasa.gov
erika.b.geier@nasa.gov

Data Management (DM) Presentation Objectives

- How are data products created for CERES?
  - Code Development
  - Delivery Process
  - Production
  - Public Release
- How do I gain access to CERES data products?
  - How do I interpret and understand the data products?
- In what condition are the Releases?
  - Availability to the Science Team vs. Public Release
- What are the future plans for DM?

CERES DM Agenda

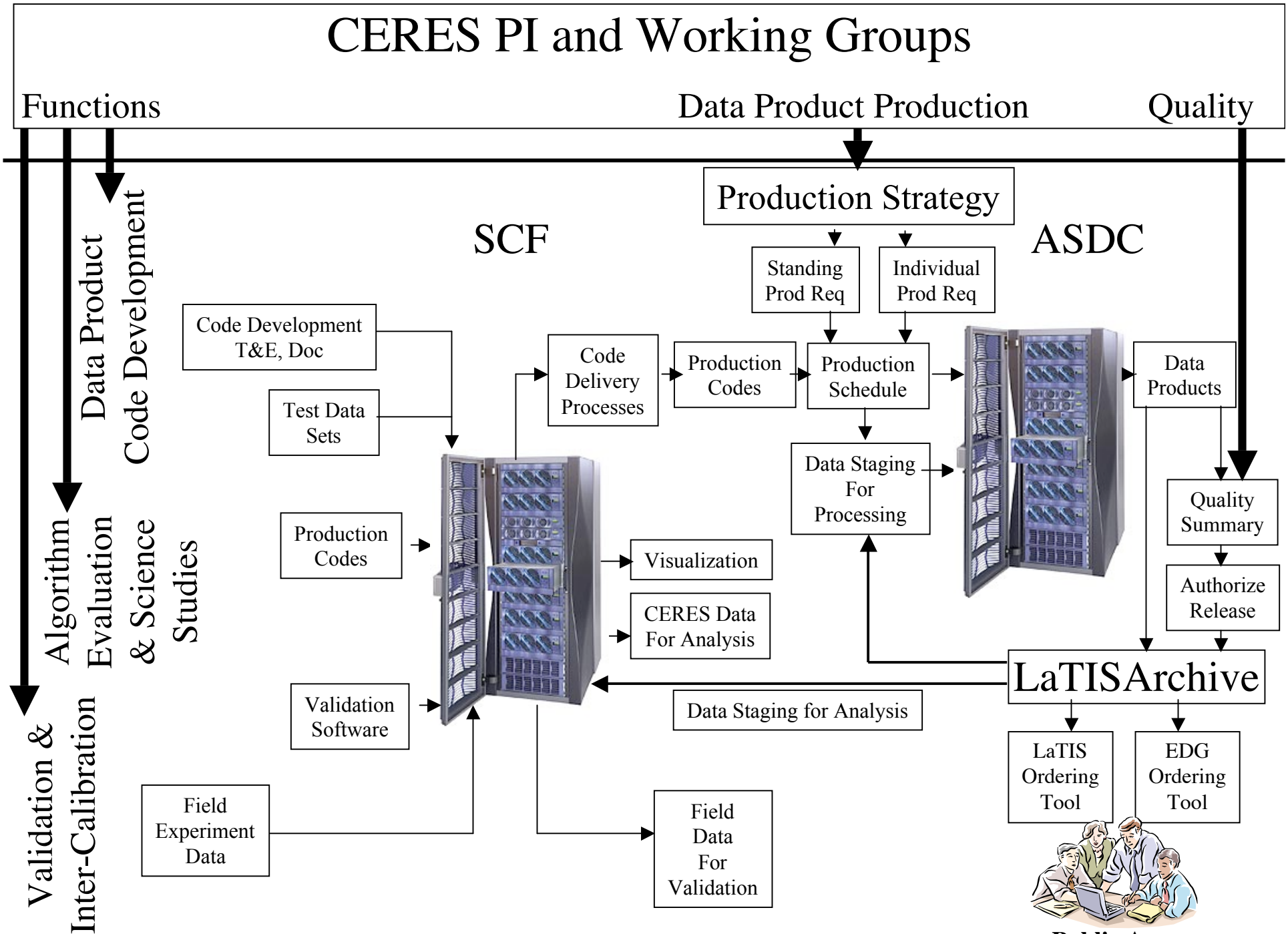
- Creating CERES Data Products
- Using CERES Data Products
  - Accessing the Data Products
  - Understanding the Data Products
    - Documentation Overview
    - Naming Conventions
    - Questions about Data Sets
- CERES Data Product Summary Tables
  - New Data Sets since November 2003
  - Data Sets expected before next Science Team Meeting
  - Older CERES Data Sets
- CERES Data Management Future Plans

Creating CERES Data Products

-Role of Data Management Team

- Provide computational support for CERES Instrument Science Team
  - Code development and testing
  - Data product evaluation
  - Validation and cross-calibration
- Translate algorithms and science codes into production code
  - Deliver to ASDC and verify products match science code outputs
- Close interaction between the instrument scientists, the code developers, the ASDC, and the data products yields best quality Climate Data Records
  - Iterative approach to algorithm development and refinement
  - Close communications due to proximity
- Complex Problem
  - 11 instruments from 7 satellites combined with 4 other inputs
    - Constantly changing characteristics
  - Algorithms and Codes are evolving
  - Objective is a highly stable data product

CERES PI and Working Groups



Creating CERES Data Products

-CERES-II Computational Environment

- Science Computing Facility
  - SGI-3800 compute servers linked to ASDC configuration
  - UNIX, Mac, PC workstations
  - Data Storage System - NOT THE PUBLIC SYSTEM
    - Working storage, temporary storage, non-public archives
    - i.e., Trial runs for comparison
  - Support infrastructure
    - Network Services
    - System Administration and User Support
    - Computer Security focused on enabling access (travel, field campaigns)
  - Software Engineering support
    - Enforce agreement on code producibility & supportability
    - Configuration Management of codes and documentation
- Atmospheric Sciences Data Center
  - Further details from ASDC Presentation
- Data management participants on Subsystem Teams

Creating CERES Data Products

-CERES Code Development

- Software Engineering
  - CERES Documentation provides guidance
    - Principles, Practices and Procedures
  - Permit Subsystem Leads independence in implementing algorithms while ensuring it can be used in production
- Current Code is only “certified” on SGI3800
  - Changes to OS, compilers, etc require testing to ensure producing the same data products
  - libraries, toolkits, input files

Creating CERES Data Products

-Delivery of Codes to ASDC

- Codes are bundled with auxiliary files, transferred to ASDC
  - Source code, test plan, operators manual, test data
  - Tested by CERES CM to verify compliance with Software Engineering requirements
- Unbundled and compiled at ASDC by CERES CM
  - Tested and evaluated by CERES CM
    - Verify completeness of delivery and compilation
  - Subsystem Integration and Test by ASDC
    - Test for production, operator instruction, output verification
- Promote to Production
- Incorporate into Production Strategy by CERES & ASDC
  - Production Requests document configurations to be produced

Creating CERES Data Products

-Production Strategy

- Close interaction between CERES DM and ASDC
  - Written Production Requests
  - Daily phone calls to resolve questions and status
  - Bi-weekly meetings to review status
- Changes to inputs and codes
  - Data availability
  - Re-processing as algorithms are refined
  - Refined versions of input data
    - MODIS
    - Replaced ECMWF data with GMAO data as it became more accurate

Creating CERES Data Products

-Support Data Product Review and Analysis

- Analysis by Instrument Science Team
  - Principal Investigator
  - Subsystem Working Group Teams
- Validation experiments
  - Field campaigns
- Inter-calibration activities

Using CERES Data Products

-Overview

- Understanding the Data Products
  - CERES Data Product Documentation
    - Data Products Catalog
    - Collection Guides
    - Data Quality Summaries
    - Description/Abstract
  - Naming Conventions
    - Maturity of data
    - Data Sets
    - Files
- Accessing the Data Products
  - Science Team members have access to non-public data
  - Public access via Web Ordering Tool and EOS Data Gateway

Using CERES Data Products

-CERES Data Products Catalog (DPC)

- Data Products Catalog (DPC) Pages for all data sets in archive
  - Static, records data product parameters as they existed at time data set was processed
  - Appropriate version of DPC pages included in every Sample Read package
- Complete Data Products Catalog
  - Living document, only most current version of DPC pages available for each product
  - Available as complete listing or individual sets of pages grouped by data product
  - Link to DPC available from CERES On-Line Documentation page **http://asd-www.larc.nasa.gov/ceres/docs.html**

Using CERES Data Products

-CERES Collection Guides

- User's Guide for Data Product
- Living document
  - Updated as needed and time/resources allow
- Eventually will be available for every CERES data product
- Meant to be used interactively
  - Fully hyperlinked and bookmarked
  - Some guides contain fully linked Index
  - Too long to print
- Linked from CERES On-Line Documentation web page

  http://asd-www.larc.nasa.gov/ceres/docs.html
- Linked from CERES Data Sets web page

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

Using CERES Data Products

-Data Quality Summary (DQS)

- User should **always** consult Data Quality Summary (DQS) before working with data set.
- Minimum set of information necessary to understand strengths and weaknesses of data set
- Living document; most up-to-date documentation
- When updates are made, e-mail is sent to all who have ordered data set
- DQS may contain helpful hints about data set
  - Consult these first when encountering a peculiarity
- Linked from CERES Data Sets web page

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

Using CERES Data Products

-CERES Description/Abstract

- Eventually will be available for every CERES data product
- Living document, update expected whenever a data set is released to the public
- Contents include
  - Data Set Description
  - Summary of Changes
  - Examples of Data
  - References
  - Contact Information
  - Acknowledgement
- Summary of Changes of particular interest
  - Typically collection of 3 tables, one for each satellite
  - Briefly lists changes and user impacts which have been made since last entry
  - Recorded by sampling-strategy, configuration code(s)
  - Includes rough date when data first available
  - Latest changes always at top; unless otherwise noted, previously recorded changes have not been undone
- Linked from CERES Data Sets web page

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

Using CERES Data Products

-Production Strategy Versions

- **Validation Data Sets**
  - Limited production run to confirm data product meets science specifications
  - Available only to Working Groups
  - Removed from ASDC upon approval to start production
- **Beta Data Sets**
  - Not of publication quality
  - May be removed from archive at any time,
    - Superceded by Edition release
  - Not all Beta data sets may be made public
  - Examples: Beta1, Beta2
- **Edition Data Sets**
  - Validated data set
  - May be used for publications
  - Will remain publicly available and in archive, even after later Editions supercede it
  - Not all data products will have an Edition1 data set
  - Examples: Edition1, Edition2B

Using CERES Data Products

-CERES Data Set Naming Convention

CER_Product-ID_Sampling-Strategy_Production-Strategy

- Example: CER_ES8_Terra-FM1_Edition2
  - CER is investigation designation for CERES
  - Product-ID specifies the data product (BDS, ES8, ES4, ES9, SSF, SFC, SRBAVG, CRS, FSW, SYN, AVG or ZAVG) Note: Product-ID may be followed by a number if size requires multiple files
  - Sampling-Strategy specifies CERES instrument(s) and, when applicable, platform and/or imager (Examples: Terra-FM2, TRMM-PFM-VIRS, Terra-FM2-MODIS, PFM+FM1+FM2)
  - Production-Strategy specifies the version (Examples: Beta1, Edition1, Edition2B)
- Within name, underscores designate change in type of information, dashes are means of separating list of items that belong together
- Preferred shorthand for referring to data set stated in Data Quality Summary

Using CERES Data Products

-CERES File Naming Conventions

CER_Product-ID_Sampling-Strategy_Production-Strategy_cc#.Date

- Same as Data Set naming convention but has Configuration Code (cc#) and data date appended to the end

  (Example: CER_SSF_Terra-FM2-MODIS_Edition2A_024025.2001031623

- Configuration Code is a 6 digit number for file and software version management, all files in a data set are considered scientifically equivalent even when configuration code changes
  - CERES recommends users always order newest file (those with greatest configuration code)
  - Users may mix and match files from same data set but with different configuration codes
- Data date preceded by a "." includes 4-digit year, 2-digit month, 2-digit day (daily and hourly products only), and 2-digit hour (hourly products only) (Examples: ".2001031623", ".20010316", ".200103")
- Gridded products may have also have a zone appended directly after the date (Example: ".199804Z09")

Using CERES Data Products

-Accessing Data Products

- NASA Langley Atmospheric Sciences Data Center
- Science Team Members
  - Access to special data sets for validation, etc.
  - Special Requests to User Services
  - Questions: Contact Erika Geier or myself
- Access by the Public
  - LaTIS Web-based ordering tool
  - EOS Data Gateway

Using CERES Data Products

-Questions about CERES 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

CERES Science Data Sets

New Since November 2003

Data Set	Comments
CER_SSF_Terra-FMx-MODIS_Edition2A	• Terra Edition2 Clouds • Final Terra ADMs
CER_SFC_Terra-FMx-MODIS_Edition2A	• Edition2A SSF are input
CER_CRS_Terra-FMx-MODIS_Edition2A	• Edition2A SSF are input • Latest version of SARB algorithms • One year (2001) of crosstrack data to be processed
CER_BDS_Aqua-FMx_Edition2	• Instrument drifts removed • 6/18/02 – 8/31/03 currently available
CER_ES8_Aqua-FMx_Edition2 CER_ES4_Aqua-FMx_Edition2 CER_ES9_Aqua-FMx_Edition2	• Instrument drifts removed • 6/18/02 – 8/31/03 currently available
CER_ES4_FM1+FM3_Edition2 CER_ES4_FM1+FM4_Edition2 CER_ES9_FM1+FM3_Edition2 CER_ES9_FM1+FM4_Edition2	• Combine only Crosstrack instruments • Jul'02 – Jun'03 currently available • uses Terra Edition2 and Aqua Edition2 inputs
CER_SSF_Aqua-FMx-MODIS_Beta2	• one month, 12/02, available
CER_SFC_Aqua-FMx-MODIS_Beta2	• one month, 12/02, available • Beta2 SSFs are input

CERES Science Data Sets

Expected Before Next Science Team Meeting

Data Set	Comments	When?
CER_FSW_Terra-FMx-MODIS_Edition2A	• Use Edition2A CRS as input • One year (2001) for crosstrack instrument	Expected to start processing in Spring 2004
CER_SRBAVG_Terra-FMx-MODIS_Edition2A	• Edition2A SFC are input	Expected to start processing in Spring 2004
CER_SSF_Aqua-FMx-MODIS_Edition1A	• 2 years of data • Aqua Edition2 IES are input • Uses final Terra ADMs	Expected to start processing in Summer 2004
CER_SFC_Aqua-FMx-MODIS_Beta3	• Edition1A SSF are input • 2 years of data	Expected to start processing in Summer 2004
CER_CRS_Aqua-FMx-MODIS_Beta1	• First Aqua SARB data set • 4 seasonal months of data • Aqua Edition1A SSF are input	Expected to start processing in Fall 2004
CER_FSW_Aqua-FMx-MODIS_Beta1	• First Aqua FSW data set • 4 seasonal months of data • Aqua Beta1 CRS are input	Expected to start processing in Fall 2004
CER_SYN_TRMM-PFM-VIRS_Beta3 CER_AVG_TRMM-PFM-VIRS_Beta3 CER_ZAVG_TRMM-PFM-VIRS_Beta3	• First Synoptic SARB and TISA averaged data sets • Process data months 4/98, 7/98, 8/98	Expected to start processing Summer 2004
CER_SYN_Terra-FMx-MODIS_Beta2 CER_AVG_Terra-FMx-MODIS_Beta2 CER_ZAVG_Terra-FMx-MODIS_Beta2	• First Synoptic SARB and TISA averaged data sets for Terra • 4 data months 1/01, 4/01, 7/01, 10/01	Expected to start processing Summer/Fall 2004

Older CERES Science Data Sets

Still of Interest (Part 1)

Data Set	Comments
CER_BDS_TRMM-PFM_Edition1	• TRMM period Jan'98 – Aug'98; Mar'00
CER_BDS_TRMM-PFM_Transient-Ops2	• Transient Operations period Jan'99 – Jul'99 • Not publicly released
CER_ES8_TRMM-PFM_Edition2	• TRMM period Jan'98 – Aug'98; Mar'00
CER_ES8_TRMM-PFM_Transient-Ops2	• Transient Operations period Jan'99 – Jul'99
CER_ES4_TRMM-PFM_Edition2 CER_ES9_TRMM-PFM_Edition2	• TRMM period Jan'98 – Aug'98; Mar'00
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	• Mar'00 only month of TRMM, Terra overlap
CER_SSF_TRMM-PFM-VIRS_Edition2B	• TRMM period Jan'98 – Aug'98; Mar'00
CER_SSF_TRMM-PFM-VIRS_Edition2B-TransOps	• Transient Operations period Jan'99 – Jul'99
CER_SSF_TRMM-SIM-VIRS_Edition2-VIRSonly	• Processed only for months when CERES was off • Currently available through Jul'01
CER_SFC_TRMM-PFM-VIRS_Edition2B	• TRMM period Jan'98 – Aug'98; Mar'00
CER_SRBAVG_TRMM-PFM-VIRS_Edition2B	• TRMM period Jan'98 – Aug'98; Mar'00

Older CERES Science Data Sets

Still of Interest (Part 2)

Data Set	Comments
CER_CRS_ TRMM-PFM-VIRS_Edition2C	• TRMM period Jan'98 – Aug'98; Mar'00
CER_CRS_ TRMM-PFM-VIRS_Edition2C-TransOps	• Not publicly released • Transient Operations period Jan'99 – Jul'99
CER_FSW_ TRMM-PFM-VIRS_Edition2C	• TRMM period Jan'98 – Aug'98; Mar'00
CER_BDS_Terra-FMx_Edition2 CER_ES8_Terra-FMx_Edition2 CER_ES4_Terra-FMx_Edition2 CER_ES9_Terra-FMx_Edition2	• Terra Edition2 BDS and ERBElike available Mar'00 through Jun'03
CER_ES4_FM1+FM2_Edition2 CER_ES9_FM1+FM2_Edition2	• Available Mar'00 through Dec'02 • Production of these data sets halted • Replaced by Terra/Aqua combined crosstrack data sets (FM1+FM3, FM1+FM4 Edition2)

CERES 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
  - Contents of this presentation: michael.m.little@nasa.gov
- CERES News (e-mail)
  - Subscribe from CERES datasets webpage
  - All new public datasets are announced as soon as available
  - Mechanism for distributing important CERES information

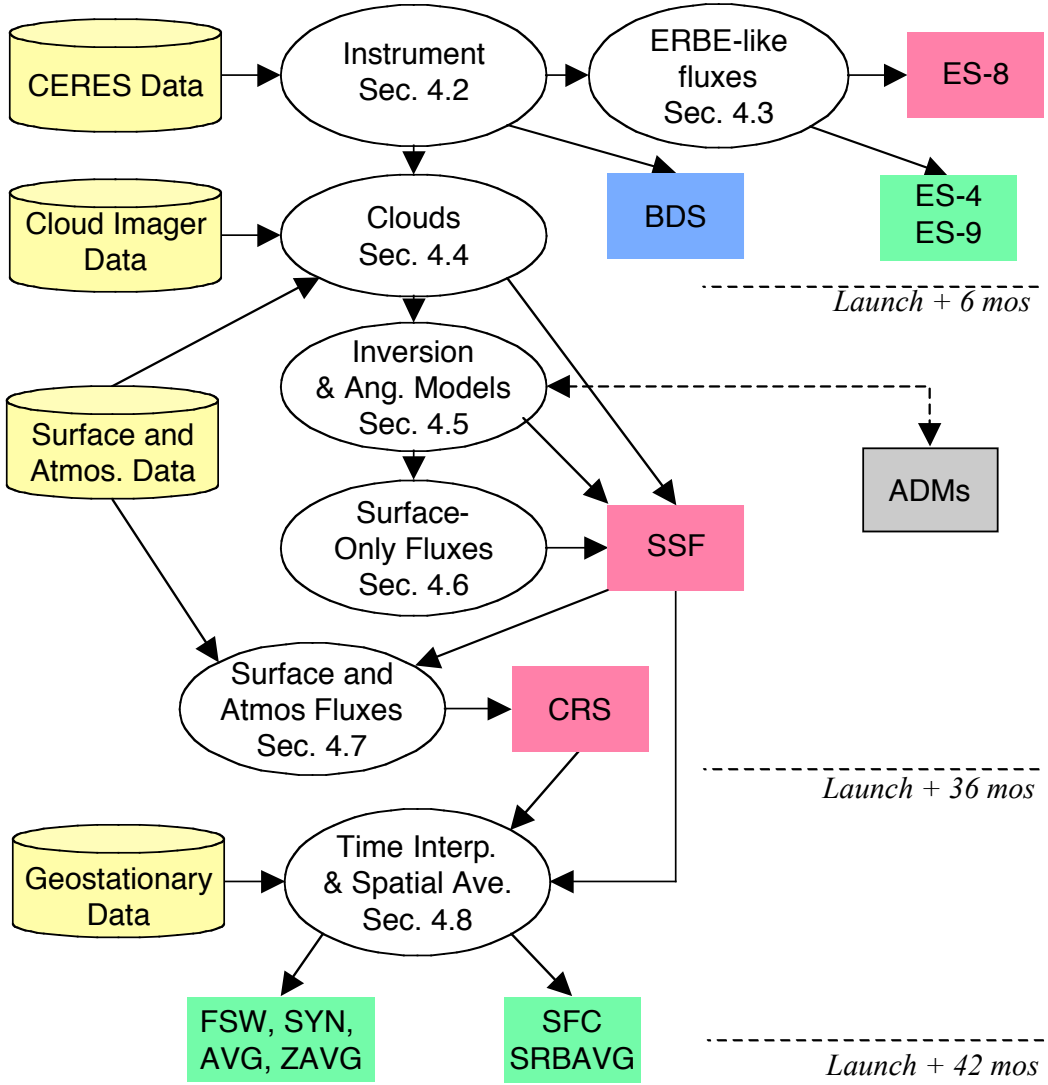
CERES Data Management

-Future Efforts

- Reduce the costs to generate CERES data products
  - Development and refinement of algorithms
  - Production of Data Products
- Increase production base options
  - Migrate to Open Source context
    - Improve validation process of stable, consistent data products on multiple computing environments
  - Support short runs and special studies
- Climate data records
  - Stability of data products
  - Calibrate, calibrate, calibrate

Backup Charts

CERES Data Products Data Flow Diagram



CERES SCF Characteristics

Feature	Aspect	Value
CERES Personnel Supported		80
Server Systems		12
Workstations		102
Bandwidth	Desktop	100 Mbps (80%)
	Between CERES Sites	50 Gigabit per second (ATM)
	to ASDC	1 Gigabit per second
External Access Controls	Current	Username/Password
	Sep03	VPN Authentication Service
Storage (May03)	Data (Programs, Data)	9 TB
	Disk Capacity	12 TB
	Archive Capacity	10 TB
	Backup Capacity	
Storage Projected Annual Growth Rate	Data	25%
	Disk	3 TB
	Archive	1 TB
	Backup	2 TB