EOSDIS Evolution at the GES DISC

From ECS to S4PA

Initial Project Review (IPR)

March 2, 2006

GES DISC Evolution Team

Gary Alcott - Gary.T.Alcott@nasa.gov
Steve Berrick - Stephen.W.Berrick@nasa.gov
Steve Kempler - Steven.J.Kempler@nasa.gov
Chris Lynnes - Christopher.S.Lynnes@nasa.gov
Bruce Vollmer - Bruce.E.Vollmer@nasa.gov

Agenda

(details on next slide)

8:30-8:45 - Introduction
8:45-10:15 - Overview of EOSIDS Evolution at the GES DISC
break
10:30-11:15 - Atmospheric Dynamics (AIRS)
11:15-11:45 - GMAO On-Line Data System (GEOS-5)
lunch
12:45-1:15 - Atmospheric Composition (Aura, SORCE)
1:15-1:45 - Terra & Ancillary Support
1:45-2:00 - Wrap-up
2:00-2:30 - Issues Recap
adjourn general session
3:00-5:00 - Executive Session

Agenda Details

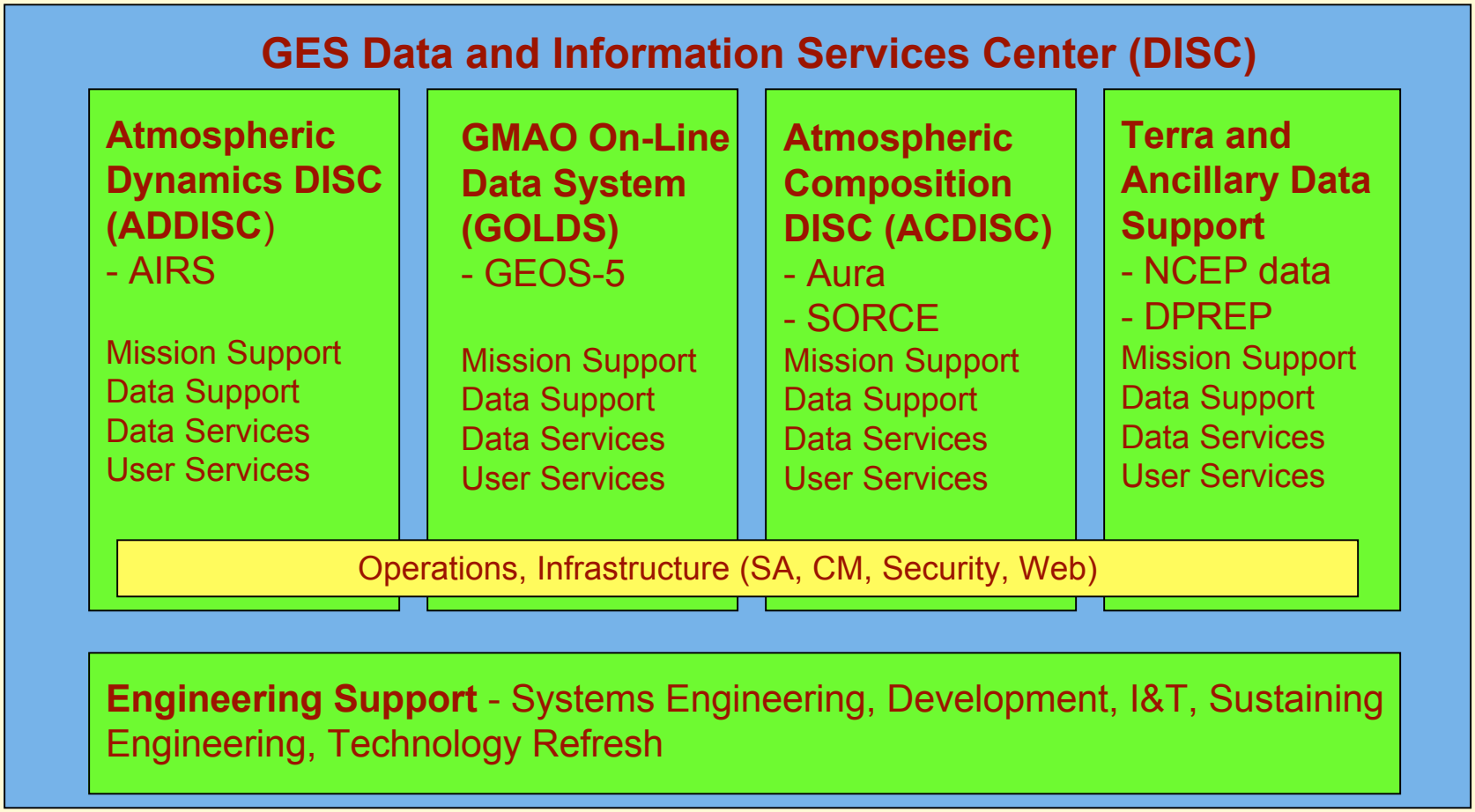
Overview
- Background and Approach
- Overview
  - GES DISC Evolution
  - Architecture
  - Requirements
- Test Plan Major
- Sustaining Engineering
- Risk Management, and Overall Risks

For Each Project
- Requirements
- Interfaces
- Architecture (including COTS)
- Assumptions and Dependencies
- Schedule
- Risks

Executive Session
- Staffing
- Budget
- Development Progress Metrics
- Security

EOSDIS Evolution at the GES DISC

GES DISC evolution is comprised of 5 projects developing 4 discipline-specific (thus, instrument specific) S4PA-based data management systems



Note: Heritage datasets were already moved to the appropriate DISC prior to this EOS data evolution

Not in Scope of This Presentation

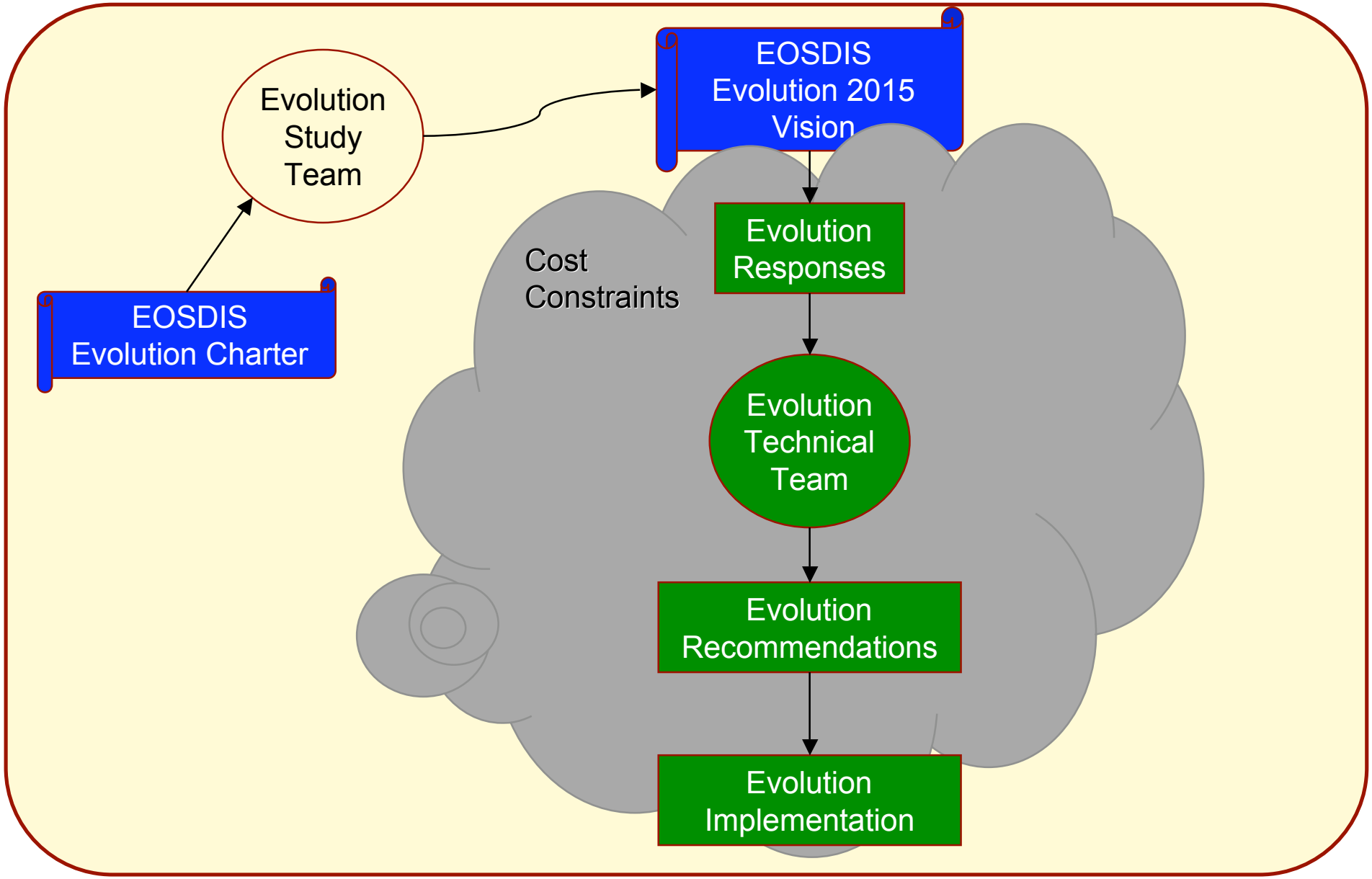
- GMAO's MERRA datasets
- Value-added functionality except for risk mitigation
- Heritage datasets (their data systems were previously evolved)
- MODIS Transition (independent effort tracked by MODIS Team)
- New Data Services

Overview of EOSIDS Evolution at the GES DISC

Initial Project Review

Chris Lynnes

Prelude


Evolution Study Team
EOSDIS Evolution 2015 Vision
EOSDIS Evolution Charter
Cost Constraints
Evolution Responses
Evolution Technical Team
Evolution Recommendations
Evolution Implementation

EOSDIS 2015 Vision (Part 1)

- NASA's research communities have access to all EOS data through services at least as rich as any contemporary science information system, for example:
  - **Data access latency is no longer an impediment**
  - The physical location of data storage is irrelevant
  - **Finding data is based on common search engines (e.g., Google2015)**
  - **Services are primarily invoked by machine-to-machine interfaces**
  - Multiple data and metadata streams can be seamlessly combined
  - **Custom processing (e.g., subsetting, averaging, reprojection) provides only the data needed, the way they are needed**
  - **Open interfaces** and best practice **standard protocols** are universally employed
- The research and value-added provider communities use EOS data interoperably with any other relevant data sources (e.g., NPOESS, METOP, GPM, numerical models, in situ systems) and systems (e.g., Global Earth Observation System of Systems).

EOSDIS 2015 Vision (Part 2)

- The EOS archive holdings are regularly peer reviewed for scientific merit:
  - Procedures for such reviews have been developed and tested over a decade
  - Derived products that are not deemed scientifically useful are phased out.
- Mechanisms to collect and preserve the pedigree of derived data products are readily available.
- Processing and data are mobile: processing can be moved to data and/or data can be moved to processing.
- NASA data systems have evolved into components that allow **fine-grained control over cost drivers.**
- Expert knowledge is readily accessible to enable researchers to understand and use the data.
- Community feedback directly to those responsible for a given system element is readily available.

Basic Approach

- *Get all the data online*
  - Eliminates acces latency
  - Enables services, machine-to-machine access, access via standard protocols
- Use dedicated archives for different measurements
  - Enables measurement (mission)-specific engineering
  - Reduces risks
  - Enables fine-grained cost control
- Reuse proven software (S4PA, S4PM)

Assumptions / Constraints

1. ECS Roll-off by 21 Dec 2007 (ESDIS)
2. Minimize changes to existing interfaces (EDOS, instrument teams)
3. Stay within current cost envelope
4. Enhance functionality and performance wherever possible within above constraints

This is not just a cost reduction exercise

What Needs to Happen for ECS Roll-Off?

- Replace functionality
  - Ingest Data
  - Store / Manage Data
  - Search Metadata
  - Access / Distribute Data
- Migrate data and responsibility out of ECS
  - MODIS ⇒ AADS
  - AIRS ⇒ S4PA
  - HIRDLS ⇒ S4PA
  - MLS ⇒ S4PA
  - OMI (+TOMS) ⇒ S4PA
  - SORCE ⇒ S4PA
  - UARS ⇒ S4PA
  - Ancillary ⇒ S4PA
- Transition interfaces
  - EDOS
  - EMOS
  - AIRS SCF
  - GMAO
  - HIRDLS
  - MLS
  - OMI
  - SORCE
  - TES
  - ECHO
  - ASTER
  - Ancillary Data
  - MODIS (covered by AADS transition plan)

Before and After: Major Changes

Before	After
One software and hardware system (ECS) for all datatypes & missions	**One software baseline**, dedicated hardware systems for missions/measurements
Data archived on tape	Data archived on disk (offline tape backup)
Users place orders for data	Users download data directly
MODIS data @ GDAAC	Value-added MODIS products only

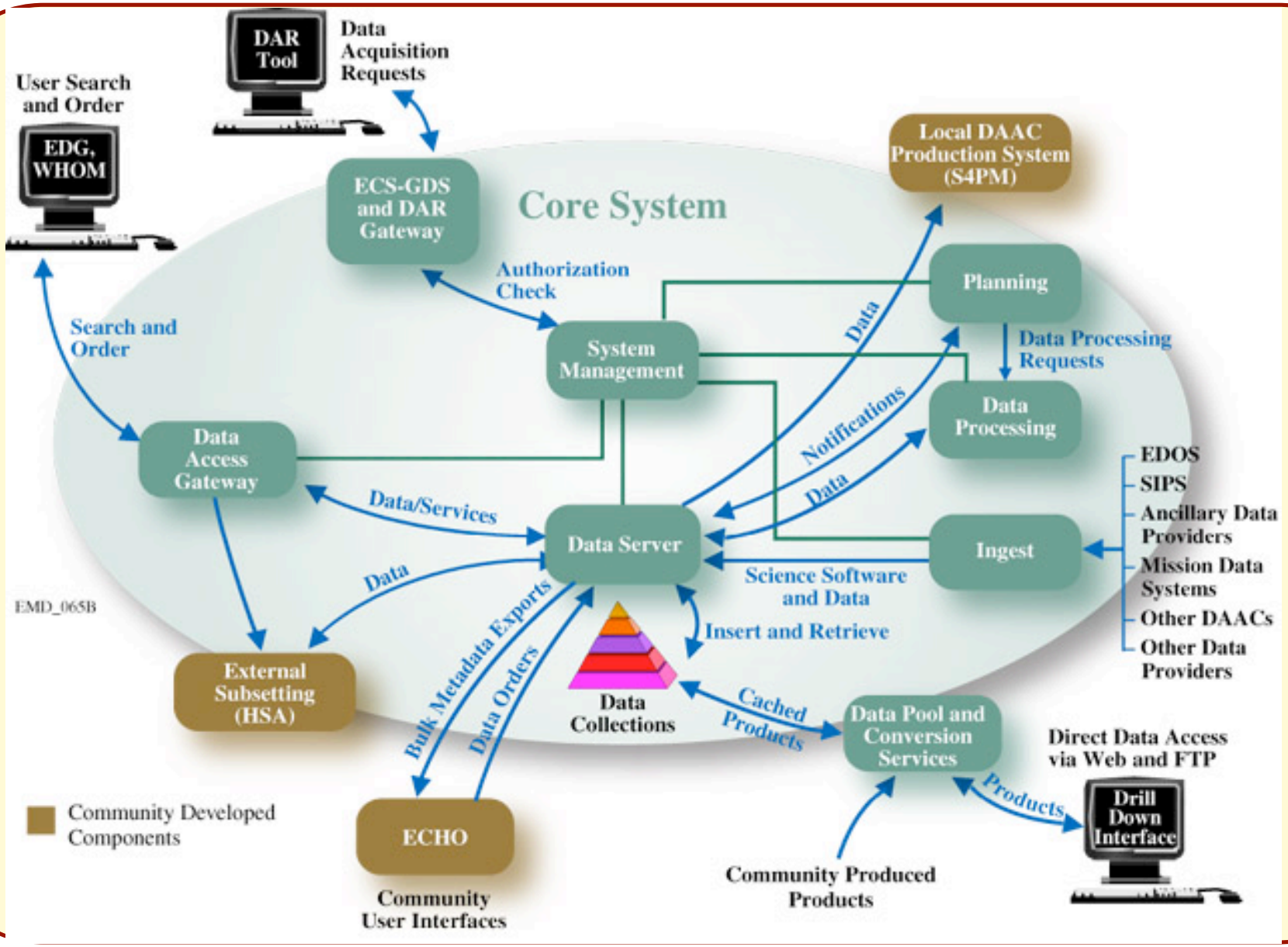
Before and After: Unchanged

- Interfaces to science teams and data producers
  - EDOS & EMOS Ingest
  - SIPS Ingest
  - ASTER Email Gateway
  - Subscription-based push
- S4PM Production
- Metadata publication to ECHO
- Metadata model
- Current data support and services

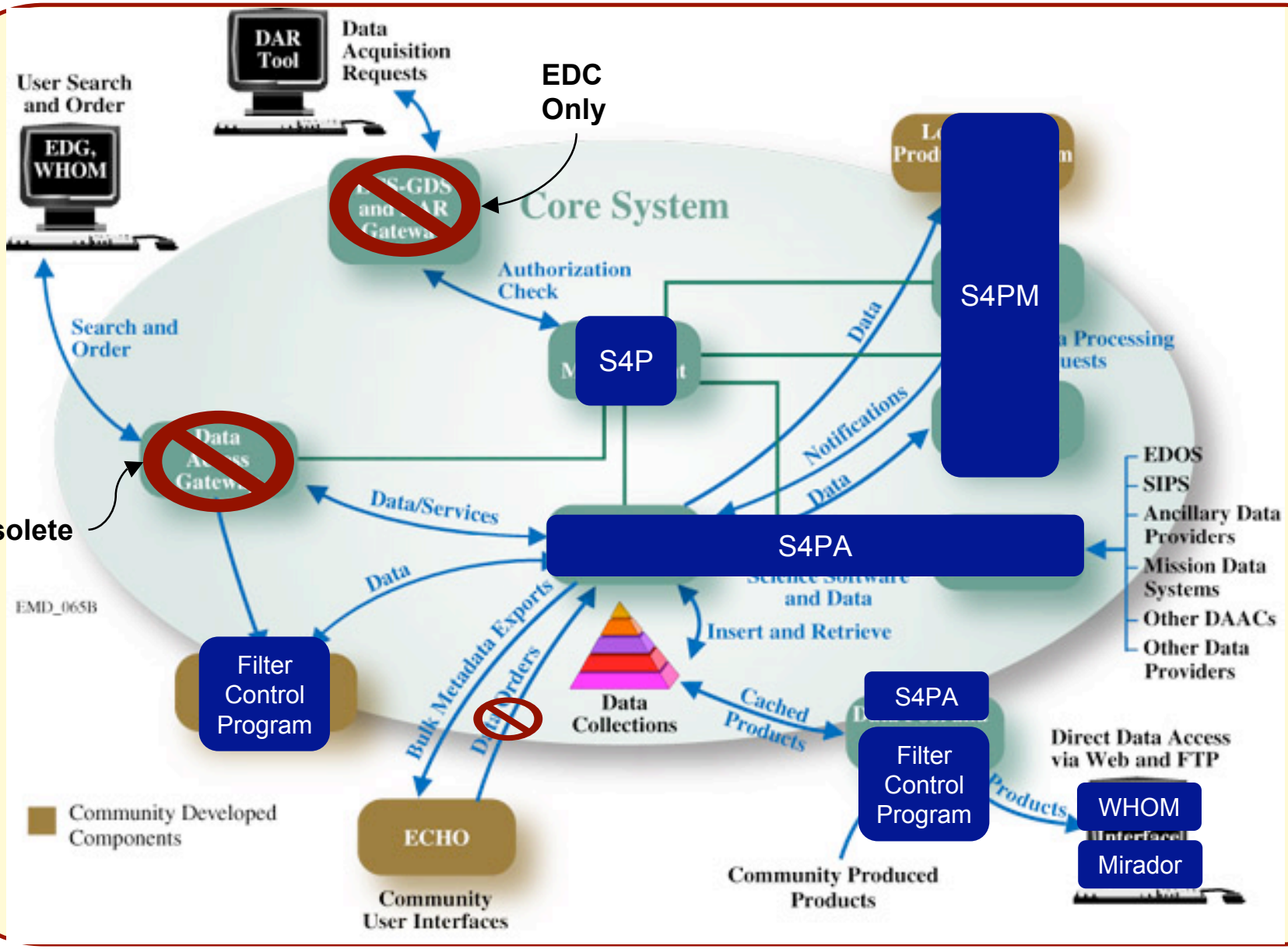
The Upside of Evolution

- Faster
  - Get data faster because it's online
  - Add services and features faster because system is simpler
  - Run reprocessing faster
- Better
  - Support machine-level applications and third party services
  - Support data mining
  - Check data integrity continually
- Cheaper
  - Need fewer operators
  - Pay less in COTS maintenance (e.g., silos)
  - Spend less in sustaining engineering

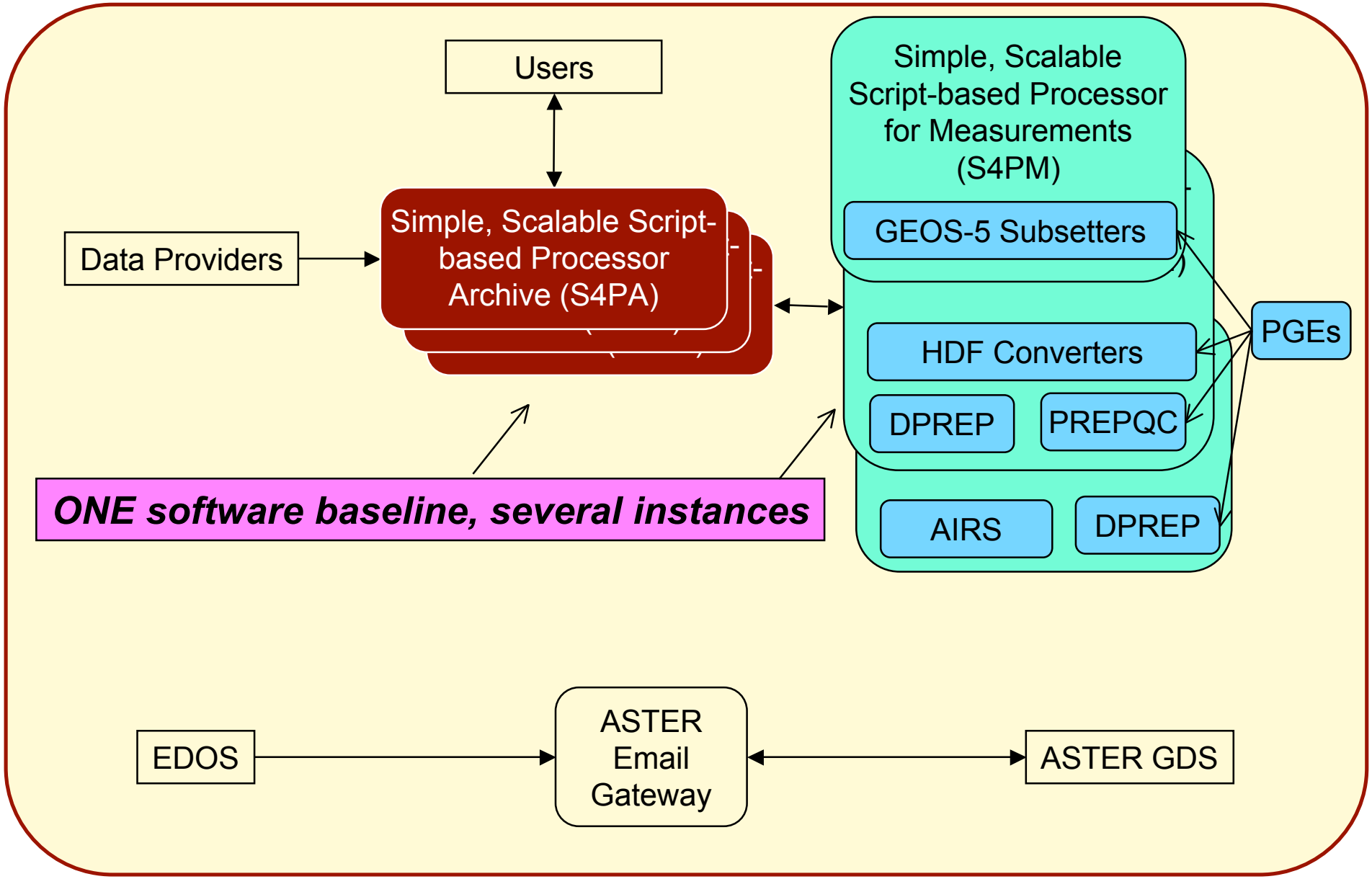
ECS Architecture Now



Architecture After



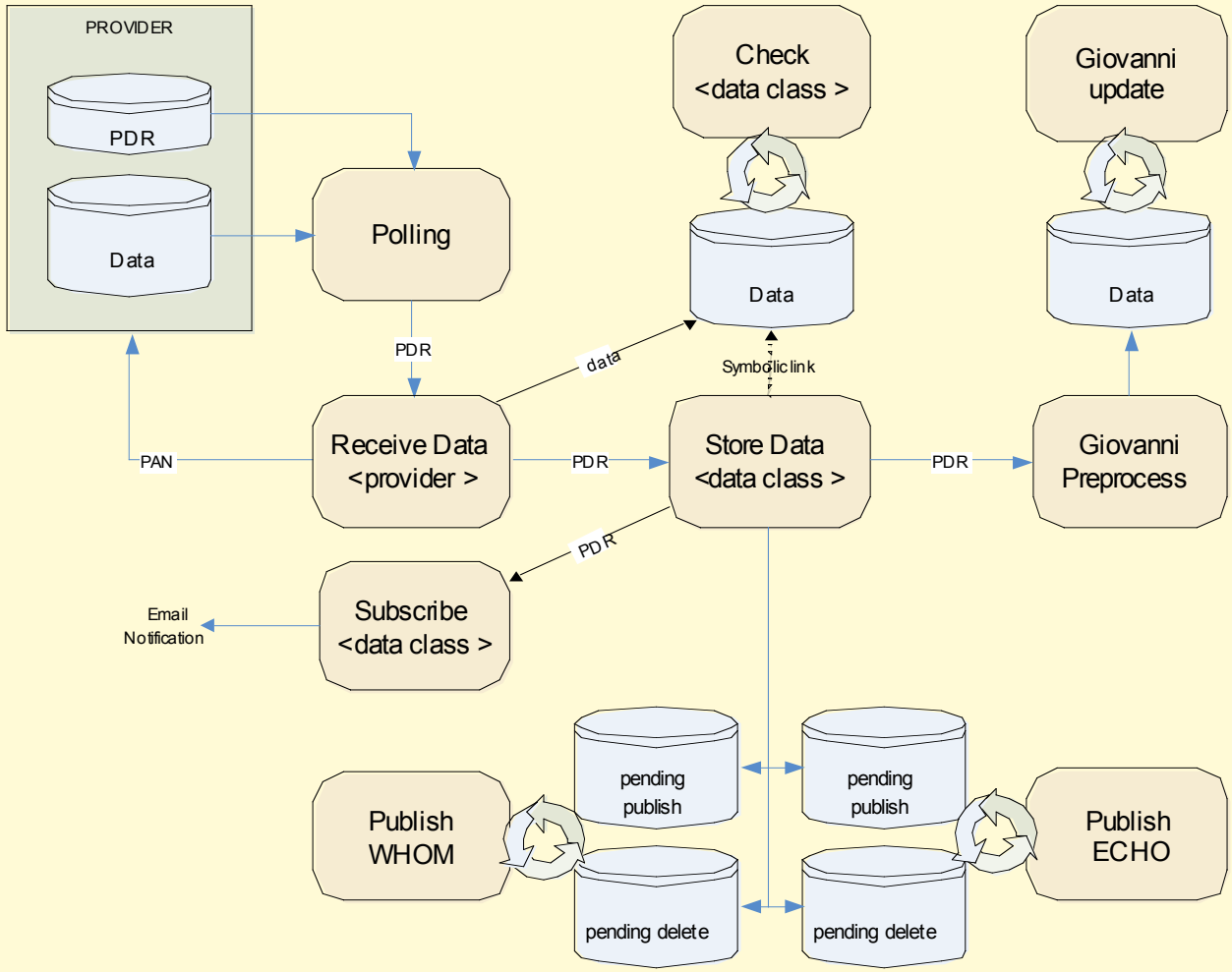
Software Architecture



What is the S4PA Software System?

- A simplified software system to automate ingest and data management for *online data*
- Based on successful S4P kernel
  - Operating since 2001 as part of S4PM
  - Reused for several processing systems
  - Implements a factory assembly-line paradigm (or DFD)
    - "Stations" = directories with configured scripts
    - Scripts process input work orders, send output work orders to downstream stations Key Characteristics
- S4PA
  - Currently supporting V0 (since 2004) and TRMM data (since Dec 2005)
  - Written in Perl
  - Compact: ~20 KSLOC

S4PA Data Flow



Current S4PA Features

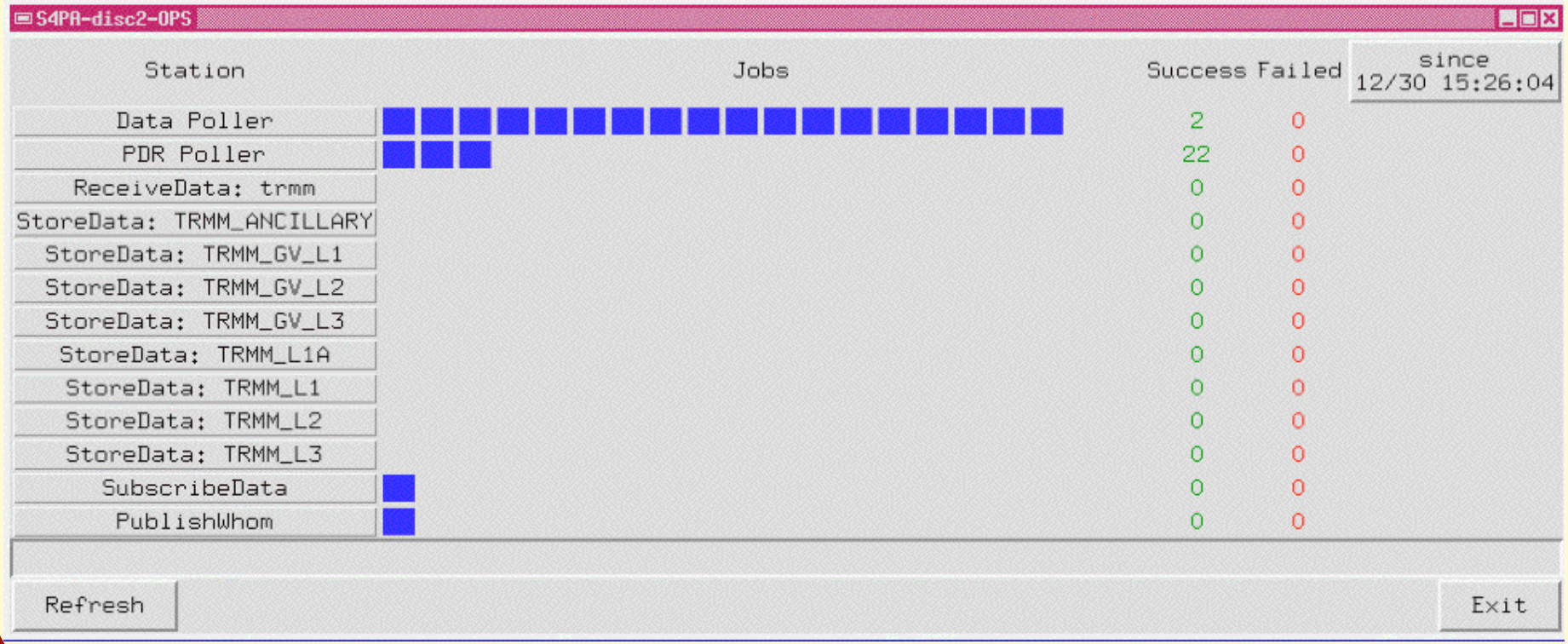
- Ingest
  - PDR / PAN
  - FTP / SCP
- Data Management
  - Metadata publication to WHOM
  - Integrity and completeness checking
  - Data replacement
  - Basic data access restrictions
  - Backup / Restore
  - Preprocessing for Giovanni
- Distribution
  - Basic FTP and HTTP access
  - Subscription notification

Main S4PA Enhancements

- ODL Metadata Ingest
- Push Subscriptions w/DN over FTP and SCP Multiple Version Support
- EDOS and EMOS Ingest
- Metadata publication to ECHO and Mirador
- Interface to S4PM
- EDGRS ingest/archive statistics reporting
- Checksum on ingest (Unix, MD5)

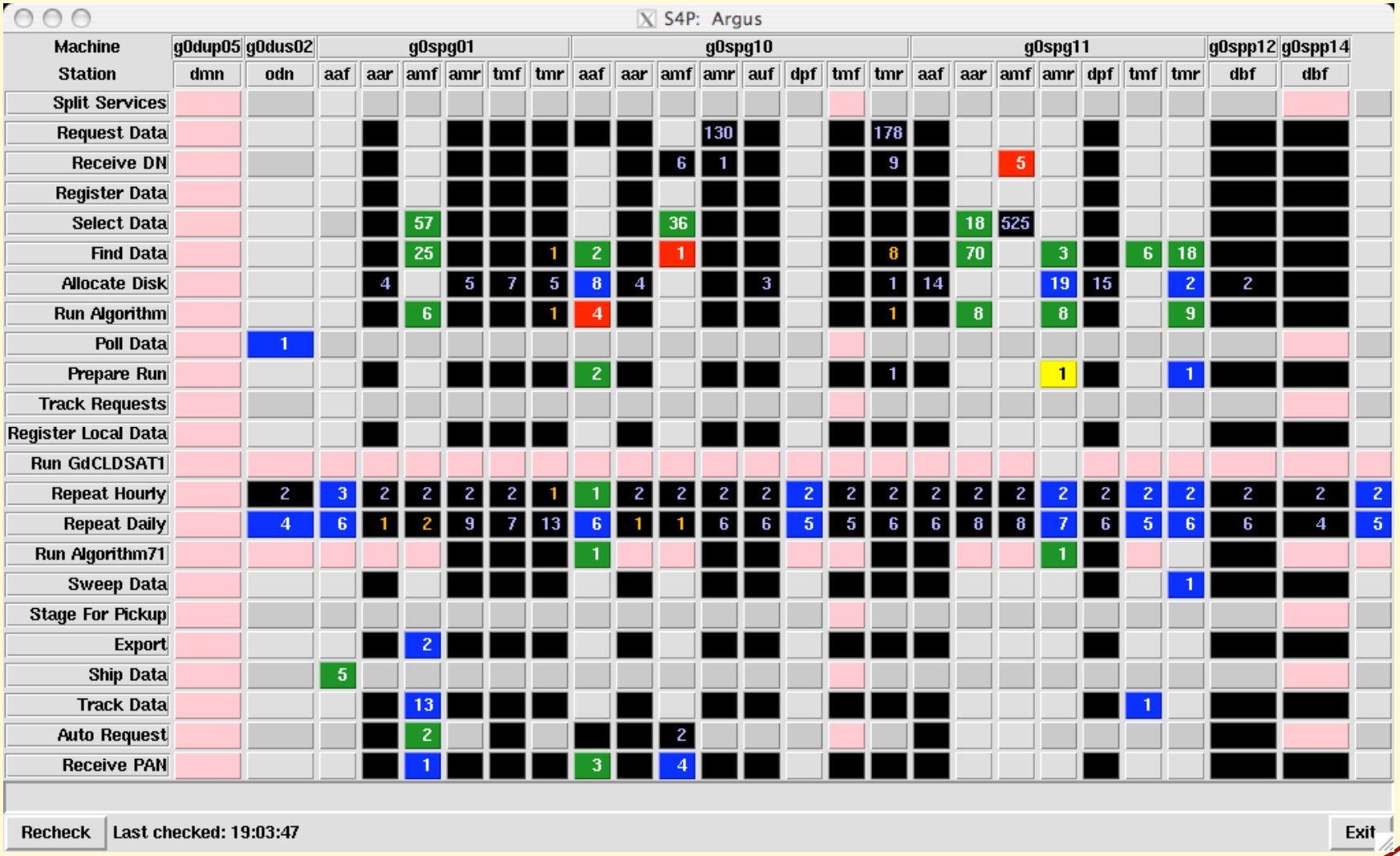
Operations Monitoring: tkstat

- tkstat: monitors S4PM or S4PA instance
  - Drilldown for troubleshooting
  - Configurable failure handlers and GUIs

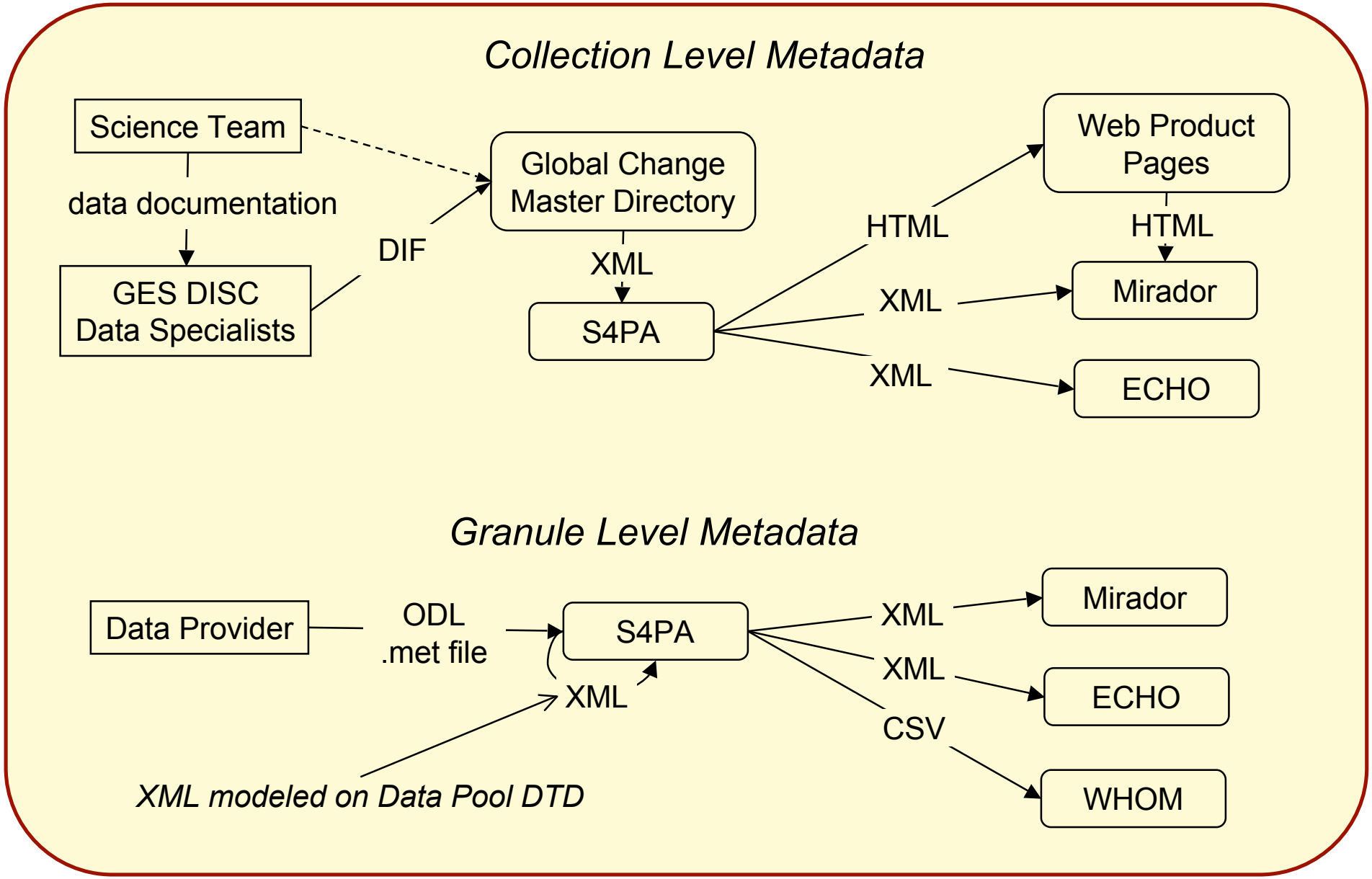


Operations Monitoring: tkargus

- monitors multiple S4PM or S4PA instances



S4PA Metadata Architecture



Minimum Required Metadata

- XML / ODL Metadata Files (S4PA Metadata Model)
  - Backward-compatible with ECS metadata model
  - Data Management Info
    - Dataset (ShortName) and VersionID
    - BeginningDateTime
    - PSAs to identify replacements (optional)
  - Additional Info for Search
    - EndingDateTime
    - Geographic and/or Orbital info if appropriate
    - Quality or product-specific attributes if appropriate
- Metadata within file
  - All metadata needed to identify, understand and use that data file

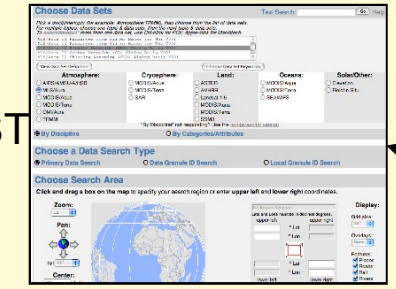
```
XML Schema:  http://disc.gsfc.nasa.gov/xsd/s4pa/S4paGranule.xsd
```

Search (Before and After)

EOSDIS-Wide

DAAC-Unique

WIST



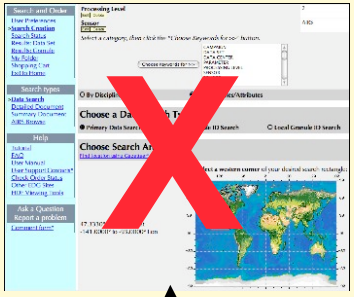
S4PA

metadata publication

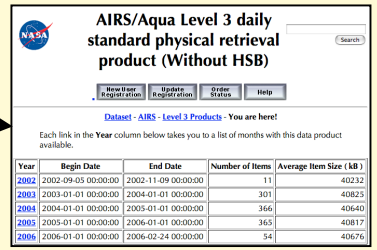
metadata publication

metadata publication

EDG



EOS Clearinghouse



WHOM

Gateway

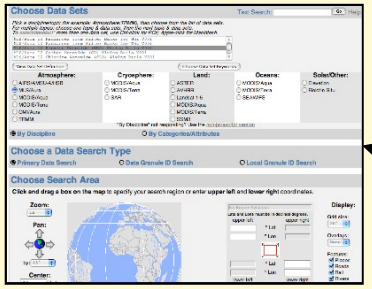
ECS

Search After

EOSDIS-Wide

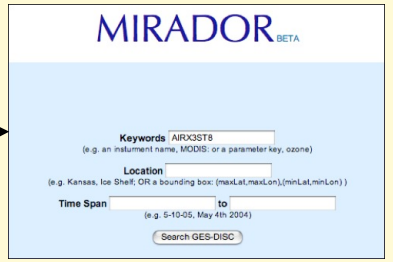
DAAC-Unique

WIST



S4PA

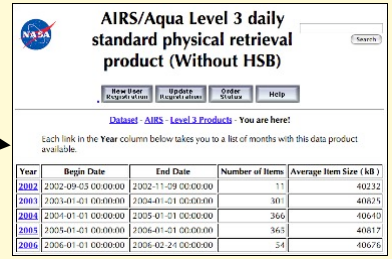
metadata publication



metadata publication

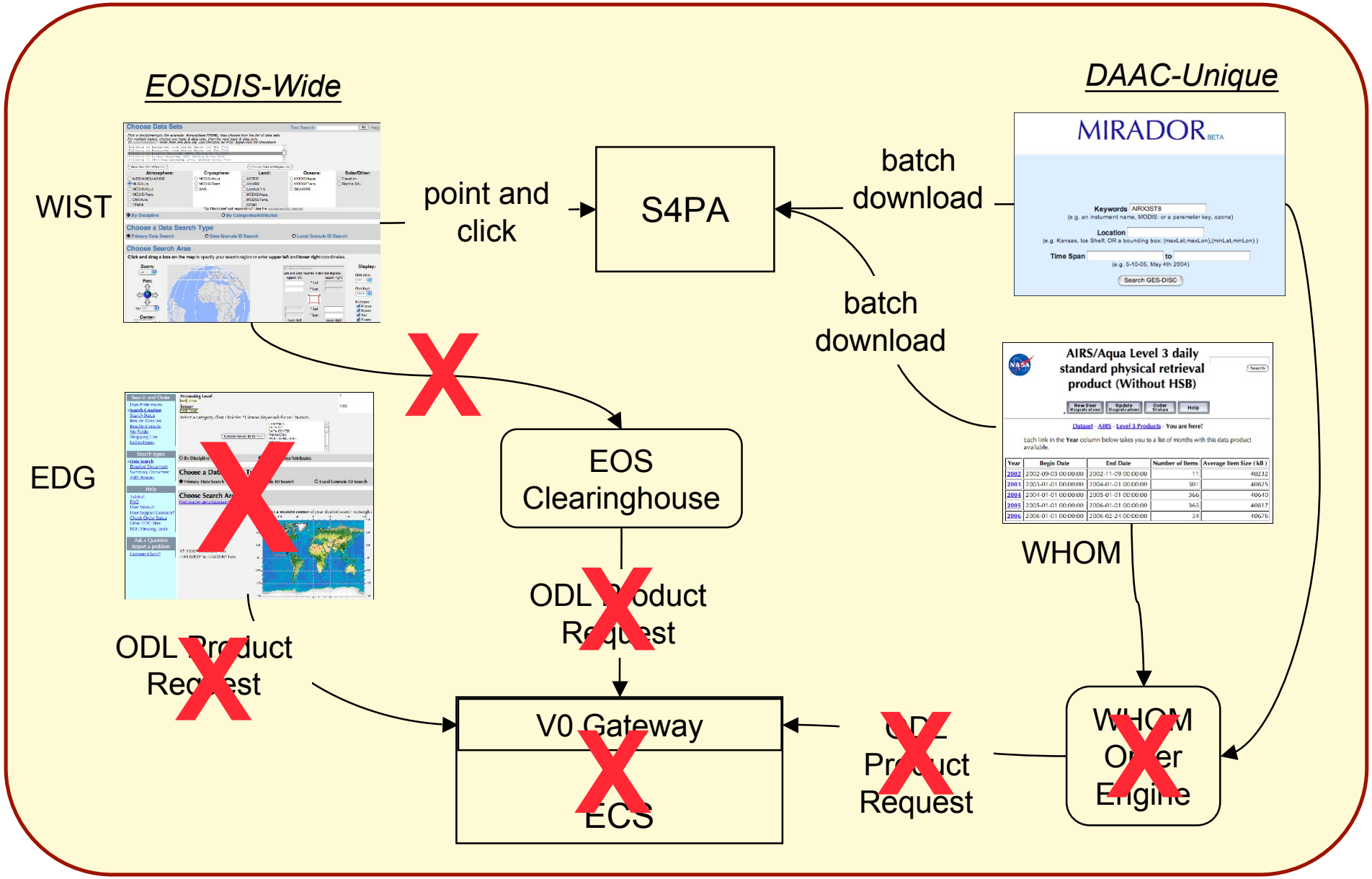
metadata publication

EOS Clearinghouse



WHOM

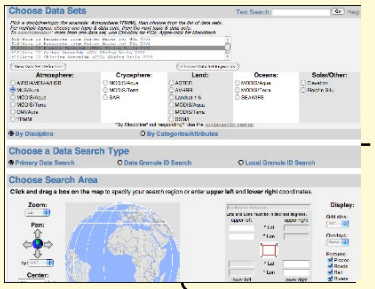
Data Access (Before and After)



Data Access - After

EOSDIS-Wide

DAAC-Unique

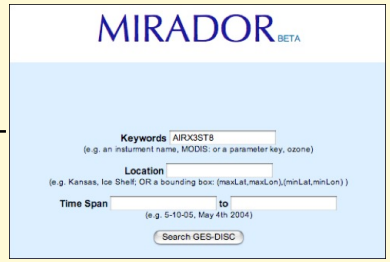


WIST

batch download ?

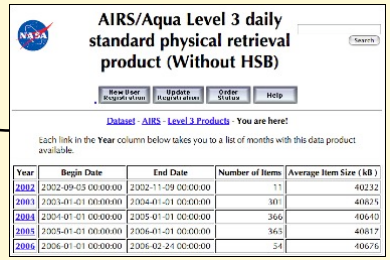
S4PA

batch download



point & click

batch download



WHOM

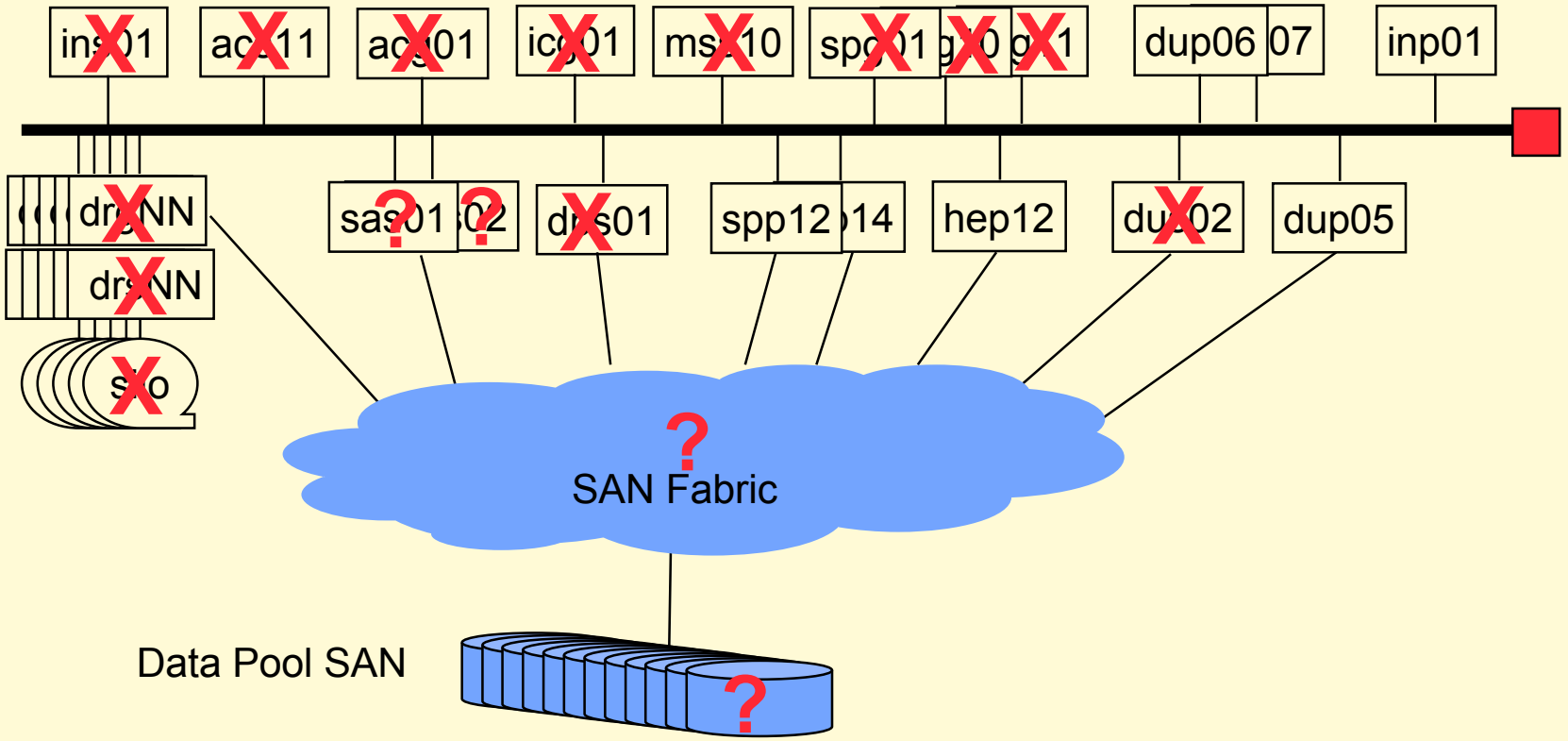
User Scenario (Mirador example)

1. User enters search criteria
2. Mirador returns datasets and estimated hits for specified time and spatial constraints
3. User selects datasets (or drills down to granules)
4. User adds selections to Shopping Cart
5. User goes to Checkout
6. User selects “Batch Download for Data and XML”
7. Browser returns batch download script
8. User executes batch script from local workstation

Transition of Current Services

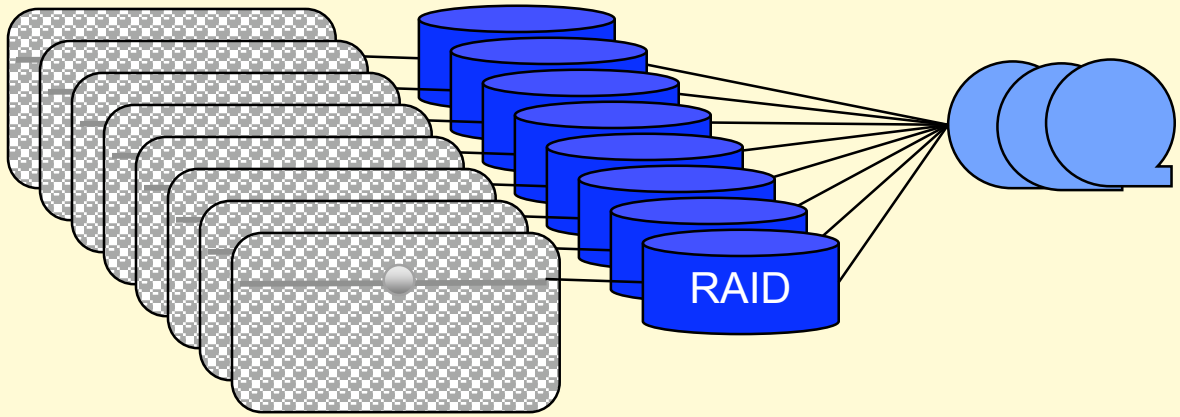
- Filter Control Program (FCP)
  - Provides simple value-added services on-the-fly, e.g., subsetting
  - Runs as WU-FTP filter program
    - Already ported to Linux
  - Will take over existing on-demand subsetting in S4PM
- Giovanni
  - Already running on Atmospheric Composition DISC (Linux) and other S4PA instances
- S4PM-DME
  - Being integrated with S4PA for TRMM forward stream
- OPeNDAP

Current Hardware: Before & After



Hardware - After (Conceptual Architecture)

Multiple Instances of Direct Attached Storage (with Tape Backup)



- Archives
  - Multiple instances of Linux systems
  - Direct attached RAID storage
  - Offline tape backup
- Processing
  - Multiple instances of Linux systems with working disk storage

Hardware Architecture Pros and Cons

- Pros
  - Simple
  - Achievable on short time scale
  - Relatively Affordable
  - Supports direct data access (cf. nearline tape)
    - Opens up data services potential
  - Compartmentalizes outages and risk (cf. ECS, SAN)
- Cons
  - Hardware provisioning takes more planning (cf. SAN)
    - May occasionally need to rehost data
  - Data inter-use requires network transfers (cf. SAN)
    - Affects value-added services only
    - Transfers can be limited by data organization, replication and reduction services (e.g. OPeNDAP)

COTS (Before and After)

Before

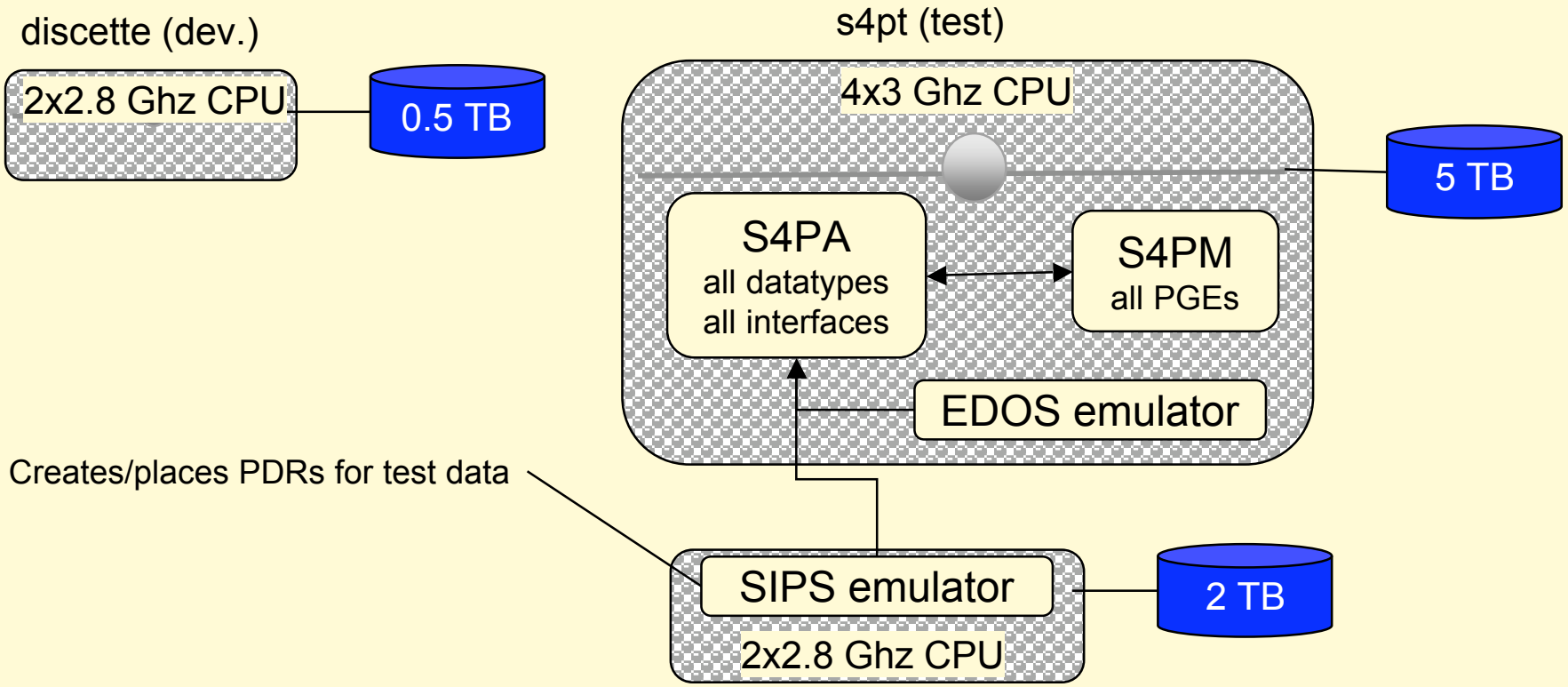
- Perl
- Sybase
- AMASS
- StorNext
- Remedy
- DDTS
- What's Up
- P.

After

- Perl
- StorNext?
- Remedy?
- DDTS?
- P.?

Test Approach

1. Unit testing
2. S4PA system testing
3. Pre-release integration testing
4. Some performance testing



5. Pre-deployment Testing on Target Machine in Test Mode
6. Mini-MOSS on Target Machine in OPS mode

Performance Testing

- Run one day(+) of data through S4PA and S4PM systems to test:
  - Input throughput (files and volume)
  - Output throughput (files and volume)
  - Processing X rates
- Performance Test Environments
  - s4pt: if successful there, larger target systems should work fine
  - Target machines: pre-operations exercises
    - Before transition to operations
    - After all, machine is not being used for anything else*
- A current data point:
  - TRMM S4PA systems routinely ingest 30 GB/day, distribute 300 GB/day

*Aura exception to be treated later

Sustaining Engineering

- Software
  - Bug Reports: Remedy
    - Release triage by Chief Systems Engineer, S4PA design lead and Project Lead of affected sub-project
      - Urgent bugs: create CVS branch and merge to next release
      - Less urgent bugs: add to next release
      - Non-urgent bugs: plan future release
  - Testing:
    - Development machine
    - Test mode on target machine(s)
- Hardware
  - Rolling upgrades: minimize risk and downtime
- Configuration Management
  - Automated CCR program
  - CCB chaired by GES DISC manager

Risk Identification Methodology

Likelihood	Consequence				
	Trivial	*Mild*	*Moderate*	*Critical*	*Catastrophic*
High (>0.5)	4	3	2	1	1
Medium (0.1-0.5)	5	4	3	2	1
Low (0.01-0.1)	6	5	4	3	2
Unlikely (0.001-0.01)	7	6	5	4	3
Improbable (<0.001)	8	7	6	5	4

High	**1-2**
Medium	**3-4**
Low	**>4**

Trivial: Minor impairment of mission success functionality or performance.

Mild: *Mission-success* function cannot be transitioned from ECS by *21 Dec 2007*.

Moderate: *Mission-essential* function cannot be transitioned from ECS by *21 Dec 2007.*

Critical: *Mission-critical* function cannot be transitioned from ECS by *31 Dec 2007*.

Catastrophic: *Mission-critical* function cannot be transitioned from ECS by *31 Jan 2008*.

Risk Actions

- Research
  - Compute
  - Benchmark
  - Spike: quick throwaway prototype
- Mitigate
  - Replan: change schedule
  - Descope: reduce functionality
  - Workaround: develop alternative workarounds
- Watch
- Accept
  - In consultation with affected parties

Overall Risks

Likelihood	Consequence				
	Trivial	*Mild*	*Moderate*	*Critical*	*Catastrophic*
High (>0.5)					
Medium (0.1-0.5)			C		
Low (0.01-0.1)			B	A	
Unlikely (0.001-0.01)					
Improbable (<0.001)					

A. Schedule compression	Response schedule of 3 yrs reduced to 2 yrs in evolution plan	*Mitigate:* frontload schedule w/ critical items, then essential items
B. Peak demand	Removing tape archive "throttle" may increase peak demand beyond machine/network capacity	*Research*: compute distribution reduction due to MODIS move *Mitigate*: limit threads for heavy users; data services to reduce output volume; divide data among several systems
C. Cooling & power	Adding disks and Intel systems may exceed C103 air handler / power capacity	*Research*: net change in power output w/removal of V0, V1, DUE, ECS machines

Overall External Dependencies

ESDIS Networks	Support of current EMS networks Help planning IPv6 compliance Plan for future network evolution
ESDIS Project	Coordination of mini-MOSSs Facilities coordination in C103 & W30
EDOS	Modifications to EDOS routing Joint interface testing
EMOS	Joint interface testing EMOS archive for products only EMOS uses
ECHO	Joint interface testing Support for ECHO ingest debugging
GCMD	Collection-level metadata management Access to XML-based DIFs (ECHO-style)
EMD	Source code for reformatters and DPREP (+RogueWave?) Estimates of maintenance cost for SAN, Remedy, DDTS, F.

Overall S4PA Software Schedule - FY06

Feb 2006	Secure (SCP) ingest & distribution
Mar 2006	ECHO interface; ODL ingest
Apr 2006	bbFTP ingest; FTP push subscription
May 2006	Basic access restrictions; L0 data
Jun 2006	Leapsec/utcpole; multiple versions; granule replacement
Jul 2006	S4PM interface; dataset relocation
Aug 2006	Group access restrictions
Sep 2006	MD5 checksum; ECS EN emulation

Overall S4PA Software Schedule - FY07

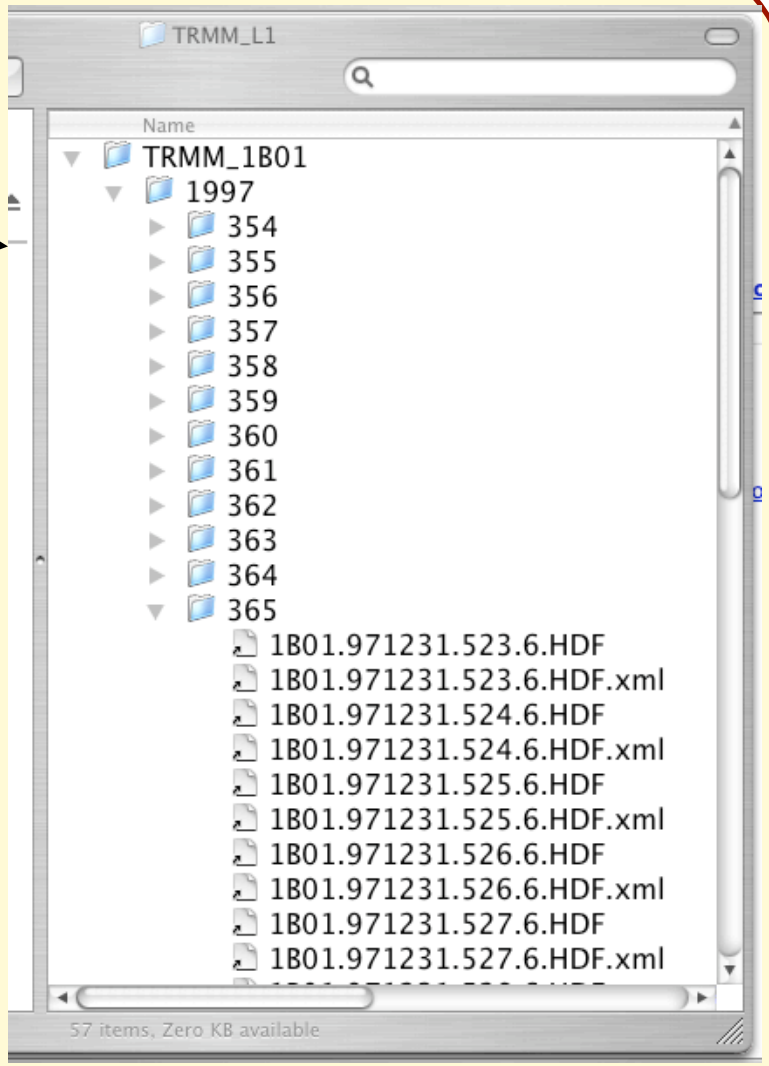
Oct 2006	PAN error handling; QA update
Nov 2006	EDGRS metrics; dot charts
Dec 2006	Batch subscription update; GCMD DIF import
Jan 2006	ECHO Browse export; PDRD handler
Feb 2007	Subscription GUI
Mar 2007	Machine-to-machine search interface
Apr 2007	ECHO OPeNDAP export
May 2007	RSS Subscriptions

TBD Items

- IPv6 Compliance by Dec 2006
  - Current ESDIS/EMD hardware mostly not compliant
- ECS domain name in network addresses
  - Transition to eosdis.nasa.gov?
  - Who? How?
- ECHO Access Restrictions
  - WIST access needs for restricted data are still uncertain
  - May be cheaper to provide search through Mirador or WHOM

TBD Items

- Machine-to-machine Gateway Replacement
  - Deterministic directory structure?
  - Web Service access to Mirador?
  - Web Service access to ECHO (DIY)?



Atmospheric Dynamics (AIRS)

Initial Project Review

Bruce Vollmer

AIRS Evolution Agenda

- Requirements
  - Data Volume
  - Data Processing
- Interfaces
- Architecture
  - Software
  - Hardware
- Test Plan & Schedule
- Assumptions & Dependencies
- Risks

AIRS Data Requirements

Processing Level	GB/day	TB/yr	Jan-06	Jan-07	Jan-08
0	12.4	4.4	15.6	20.0	24.4
1	17.4	6.2	27.9	34.1	40.3
2	8.1	2.9	13.0	15.9	18.8
3	0.05	0.02	0.08	0.1	0.11
Total	38	14	57	70	84

- Distribution Mechanisms: FTP Push/Pull (same functionality as current)
- Access Mechanisms: ECHO clients (WIST), WHOM, Mirador; restricted access to data yet to be validated
- Maintain Giovanni services for AIRS online data
- Requirements for Archiving, Distribution and User Services in EOSDIS

AIRS Requirements

- Data Processing
  - 1x forward (Mission Critical)
    - keep up with input data flows
  - 2x reprocessing (Mission Critical)
    - Modest requirement; new architecture will provide faster reprocessing rates consistent with goals of evolution and measurement based systems
    - Mission essential: reprocess in 1 year (5x)
    - Mission success: reprocess in 6 months (10x)

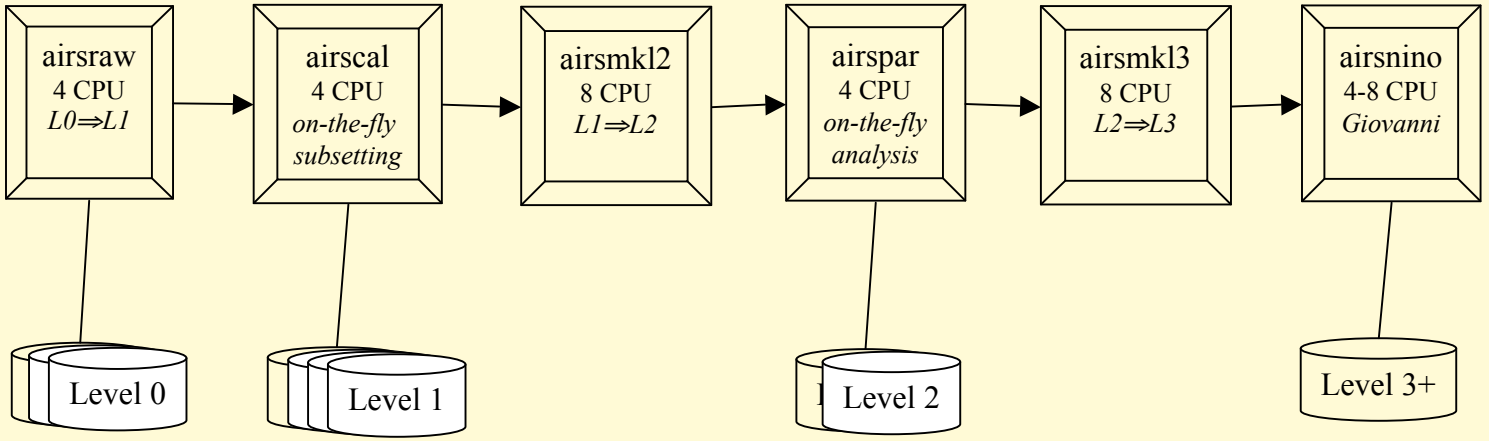
AIRS Interfaces

Interface	Data	Transfer Mechanism
EMOS	Input: att, eph, eng data	FTP Push
EDOS	Input: L0 ancillary, science data	FTP Pull
NOAA	Input: dynamic ancillary data	FTP Pull
USNO	Input: UTCPoleT, LeapsecT data	FTP Pull
AIRS SCF	Input: Algorithm Software Output: Selected L0, L1-L3	FTP Push
NSIDC	Output: Aqua DPREP output (AMSR-E)	FTP Pull
Users	Output: L1-L3, Browse	FTP Push, Pull

AIRS Architecture

- Software
  - S4PA software system with support for FTP Pull, FTP Push, secure shell distribution (if needed) and access restrictions (where appropriate)
  - S4PM software system to run AIRS algorithm software
  - On the fly subsetting algorithms (FCP)
    - Current OTF, transition ODN
  - Maintain Giovanni, OpenDAP servers

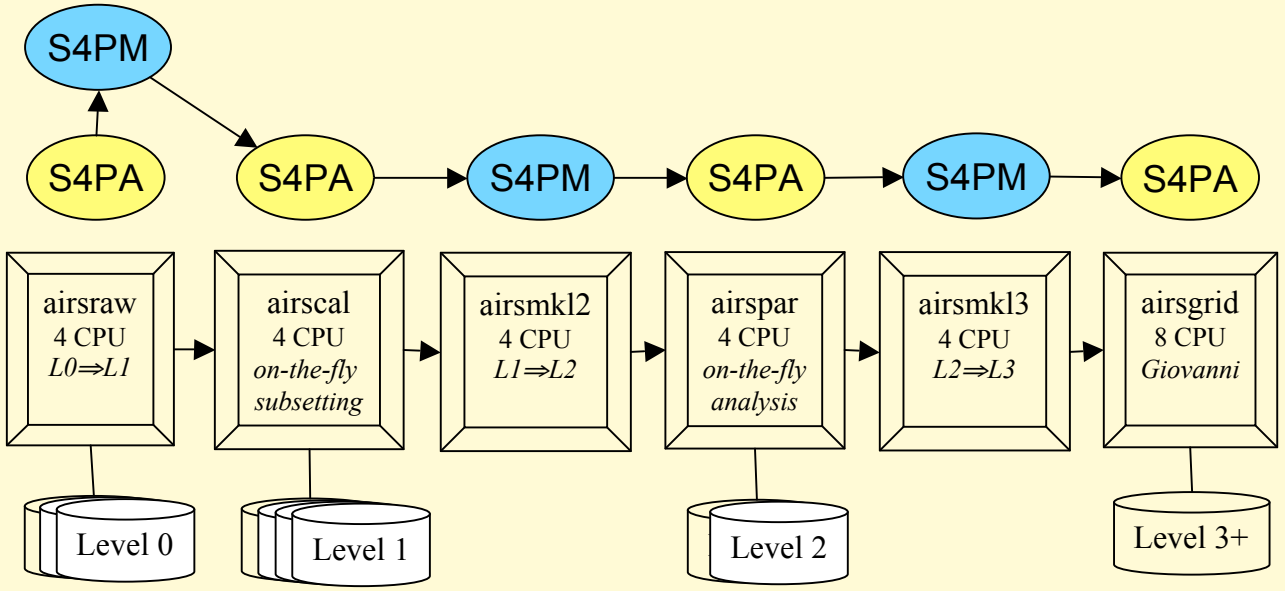
AIRS Architecture



Insulated Level-Slice Architecture

+ Optimized hardware for each task
+ Simple reprocessing scheme
+ User access segregated from processing machines
– Significant data movement (Production Network)

Note: Actual number of platforms and cpu allocations dependent on benchmarking activities



Insulated Level-Slice Architecture

+ Processing hardware insulated from user demand
+Optimized hardware for each task
+ Simple reprocessing scheme
– Maximum data movement

AIRS Assumptions and Dependencies

- Two month turnaround for hardware procurement
  - PR submission to hardware delivery
- Port of AIRS algorithm software to linux complete by start of V5 SSIT

AIRS Draft Schedule and Test Plan

Start	Finish	
1/4/2006	2/24/2006	AIRS S4PA Hardware design
2/1/2006	2/15/2006	Set up linux processing test platform
2/1/2006	4/28/2006	Local port of AIRS V4 Alg SW to linux
2/27/2006	4/28/2006	Procure Hardware
5/1/2006	5/26/2006	Install and configure hardware
6/5/2006	9/22/2006	Migrate retrospective L0 data to S4PA
	6/15/2006	AIRS V5 Algorithm SW delivery to DISC
6/19/2006	8/4/2006	AIRS V5 SSIT
7/17/2006	8/4/2006	Integrate with ECS (DPREP, Ancillary data)
7/31/2006	8/18/2006	Integrate with EDOS (L0 data)
8/7/2006	8/18/2006	End-to-end integration
8/25/2006	9/7/2006	End-to-end system testing

AIRS Draft Schedule and Test Plan

Start	Finish	
9/20/2006	9/22/2006	End-to-end test (Mini-MOSS)
9/25/2006	10/6/2006	Shadow processing C5 forward (AIRS validation processing)
10/9/2006	10/13/2006	Ramp up C5 processing/reprocessing (ECS idle)
1/8/2007	1/26/2007	Aqua DPREP, PREPQC SSIT
1/8/2007	1/26/2007	Integrate with EMOS (attitude, ephemeris)
1/29/2007	2/16/2007	End-to-end integration
2/21/2007	2/23/2007	End-to-end test (Mini-MOSS)
1/8/2007	1/26/2007	Aqua DPREP, PREPQC SSIT
	2/28/2007	AIRS Transition Readiness Review
	3/1/2007	Turn off ECS for AIRS

* Schedule to be worked with AIRS/JPL

AIRS Risks

Likelihood	Consequence				
	Trivial	*Mild*	*Moderate*	*Critical*	*Catastrophic*
High (>0.5)					
Medium (0.1 -0.5)			**B**		
Low (0.01 -0.1)		**C**	**A**		
Unlikely (0.001 -0.01)					
Improbable (<0.001)					

A. Linux port of AIRS PGEs	PGEs installed, configured and tested by start of C5 processing	*Research:* perform early porting to available test platform *Mitigate:* start C5 processing later *Workaround:* Use ECS HW and SW if delay is significant and science drives start of C5 processing
B.Availability of hardware	Mixed track record with HW vendor	*Mitigate:* same as above *Workaround:* alternate vendor
C. Throughput capacity	Target X-rates (5-10x) may not be achievable on system architecture	*Benchmark*: use available platforms to establish early benchmarks

AIRS Risks

- Port of AIRS Alg Software to linux
  - Porting has started with V4 algorithm software
  - AIRS JPL and U Wisc successfully ported for DB
- Hardware available for system use
  - Mixed track record with Western Scientific
- Throughput capacity
  - I/O, processing rates

AIRS Questions

- What is the likelihood of June 2006 delivery of AIRS PGEs to GSFC?
  - Key driver for C5 processing
- Proper triage of AIRS summary browse support
  - Currently essential (recommend move to success)
  - Primary impact on WIST

GOLDS (GEOS-5)

Initial Project Review

Steve Berrick

GOLDS Agenda

- Requirements
- Interfaces
- Architecture
  - Software
  - Hardware
- Test Plan
- Assumptions & Dependencies
- Schedule
- Risks

GOLDS: Requirements

- Data Volumes:

Processing Level	GB/day	TB/yr	Jan-06	Jan-07	Jan-08
4	9†	3.2	0	1.6‡	4.8
Total	9	3.2	0	1.6	4.8

† Assumes average compression efficiency of 30%
‡ Assumes start of flow on 1 July 2006

GOLDS: Requirements

- Data Rates:

Flow	Read/Write	Processing Rate	Data Rate (GB/day)
GEOS-5 Forward	Write	1x	9†
GEOS-5 Reprocessing	Write	16x	144
To Subscribers	Read	1x	18‡
To General users	Read	5x*	45
Total		23x	216

* Guess
† Assumes average compression efficiency of 30%
‡ Based on current GEOS-4 sizes adjusted for resolution and compression

GOLDS: Requirements

- Other Requirements:
  - Ready for start of GEOS-5 forward flow to GMAO On-Line Data System (GOLDS): 1 July 2006
  - GEOS-4 “frozen” D4C* CERES products to LATIS through Jan 2008
  - Distribution Mechanisms: FTP Push/Pull, scp (same as current)
  - Ingest Mechanism: bbFTP
  - Access Mechanisms: Support for collection and granule level access restrictions
  - Subsetting services: HIRDLS, MLS, MODIS (same as current)

GOLDS: Interfaces

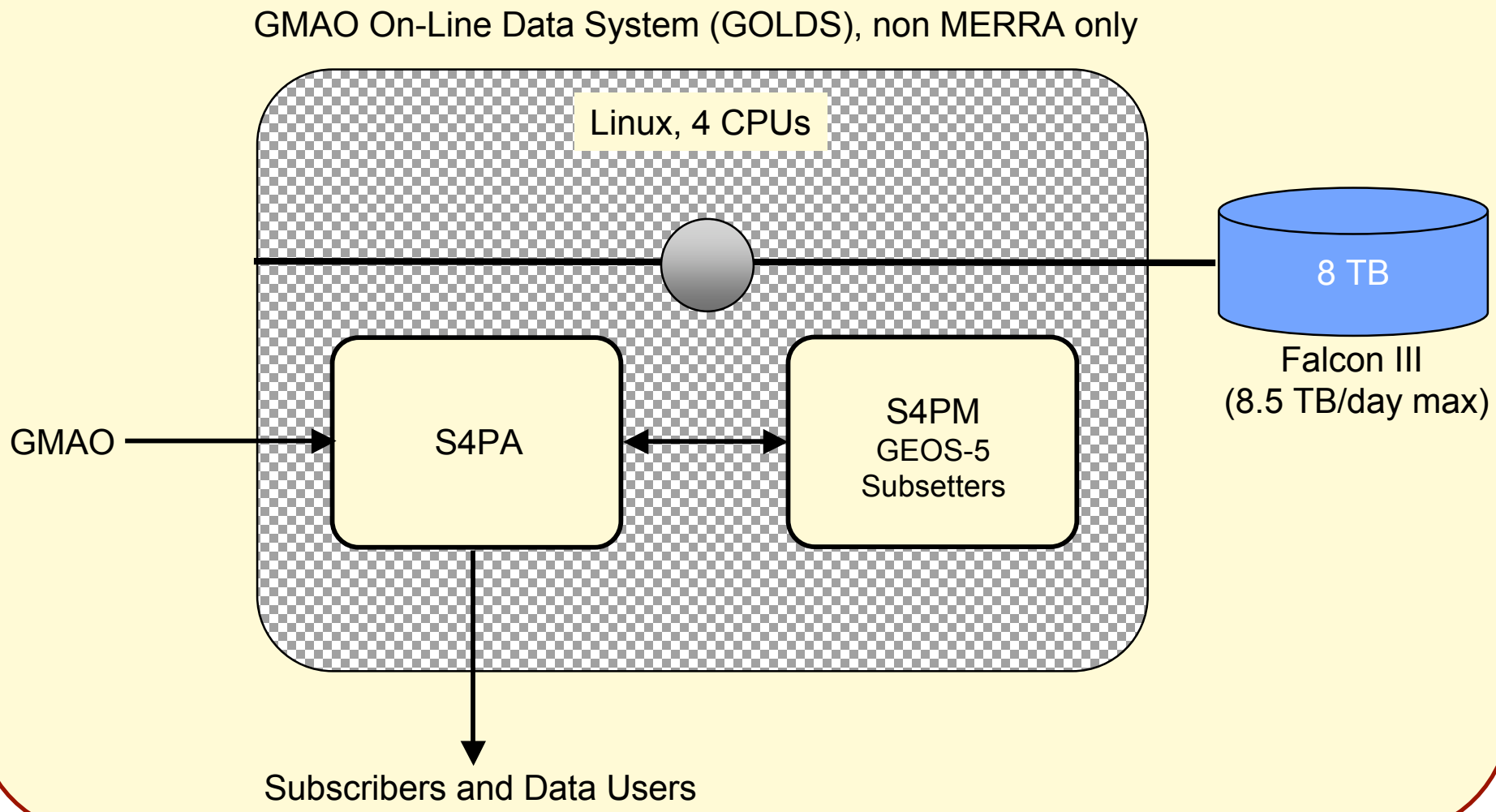
Interface	Data	Transfer Mechanism
GMAO to GOLDS	GEOS-5	bbFTP
GOLDS to LATIS at Langley	"Frozen" GEOS-4 products for CERES (through Jan 2008)	FTP Push
	GEOS-5 for non CERES LATIS	FTP Push
GOLDS to CALIPSO and others in ECS at Langley	GEOS-5	FTP Push
GOLDS to HIRDLS SIPS	GEOS-5	FTP Push
	HIRDLS Subset products	FTP Push
GOLDS to MLS SIPS	GEOS-5	scp
GOLDs to MLS SCF	GEOS-5	scp
	MLS Subset products	scp
GOLDs to MODIS	University of Montana subset products	FTP Push

GOLDS: Interfaces

- Interface changes in GOLDS over ECS:
  - New IP addresses

GOLDS: Architecture

- Hardware



GOLDS: Architecture

- Software
  - S4PA software system with support for FTP Pull, FTP Push, and secure shell (scp) distribution as well as access restrictions
  - bbFTP for ingest support
  - Subsetting algorithms
    - 2 MLS algorithms
    - 4 HIRDLS algorithms
    - 2 MODIS algorithms
  - S4PM software system in which to run the algorithms
  - Mitigation strategy software: Giovanni, OPeNDAP servers

GOLDS: Test Plan

- Provide early samples of subset products generated in S4PA test system (s4pt) to intrument teams
- System testing of interfaces and algorithms on test mode of GOLDS
  - Run “day in the life of” tests
  - Evaluate functionality and performance
- Promote into operations mode on GOLDS
- Mini-MOSS and Test Readiness Review (TRR)
- Resolve any issues resulting from TRR

GOLDS: Assumptions and Dependencies

- GEOS-5 file spec finalized by 31 Mar 2006
- GEOS-5 forward flow to GOLDS starts 1 July 2006
- GEOS-5 reprocessing flow to GOLDS starts NET 1 Dec 2006
- GEOS-5 volume is 8.87 GB/data day (assumes 30% compression efficiency)
- Subsetting software requirements (for MLS, HIRDLS, and MODIS) remains simple (no more complex than current subsetting algorithms)
- GEOS-5 sample data available by mid April 2006 (for algorithm development)
- Current GEOS-4 products continue to flow to ECS for 1.5 months after GEOS-5 starts (through 15 Aug 2006)
- CERES need for "frozen" GEOS-4 ends Jan 2008
- General user download from GOLDS, once public, is 5x

GOLDS: Schedule

05/31/06	Receive GOLDS hardware
06/13/06	Complete GOLDS software installation and setup (S4PA, S4PM, etc.)
05/01/06- 06/30/06	Subsetter algorithm porting with sample products reviewed by instrument teams
06/14/06- 06/28/06	Early testing of GOLDS interfaces in S4PA test system (s4pt)
07/03/06- 07/05/06	Testing of algorithms in S4PA test system (s4pt) with sample output reviewed by instrument teams
07/06/06- 07/07/06	Promotion of interfaces and algorithms into test mode on GOLDS
07/10/06- 07/21/06	"Day in the life of" tests in test mode on GOLDS with all interfaces

GOLDS: Schedule

07/24/06- 07/28/06	Promotion of interfaces and algorithms into operations mode on GOLDS
08/01/06- 08/04/06	Mini-MOSS testing
08/07/06	Test Readiness Review
08/08/06- 08/11/06	Resolve any issues resulting from TRR
08/14/06	ECS not needed for GEOS-4 flow (except for CERES)

GOLDS: Risks

Likelihood	Consequence				
	Trivial	*Mild*	*Moderate*	*Critical*	*Catastrophic*
High (>0.5)					
Medium (0.1-0.5)		A			
Low (0.01-0.1)			B		
Unlikely (0.001-0.01)					
Improbable (<0.001)					

A. bbFTP problems	bbFTP for GMAO to GOLDS flow could prove to be unreliable	*Research*: dig further into issues of performance and reliability *Mitigate*: we can easily switch to scp
B. More complex subsetters	Due to new bundling of parameters in GEOS-5, subsetters may become more numerous or complex	*Watch*: Subscribers have not indicated any need thus far

GOLDS

Backup Slides

GOLDS: Current GEOS-4 Subscribers

- Langley ECS
  - D4FAPMIS, D4FAXMIS, D4LAPCHM, D4LAPCLD, D4LAPMIS, D4LAPMOM, D4LAPMST, D4LAPTMP, D4LAPTRP, D4LAXCHM, D4LAXCLD, D4LAXLSM, D4LAXMIS, D4LAXSTR
  - 359 GB/year
- Langley LATIS
  - D4FAPCHM, D4LAPCHM, GMAOHIR1, GMAOHIR2, GMAOHIR3, GMAOHIR4
  - 272 GB/year
- MLS SIPS
  - D4FAPMIS, D4FAXMIS
  - 72 GB/year
- MLS SCF
  - D4CAEMLS, D4FAEMIS, D4FAXCLD, D4LAEMIS, D4LAPMIS, D4LAXCLD, D4LAXMIS, GMAOMLS1, GMAOMLS2
  - 911 GB/year
- HIRDLS
  - D4FAPCHM, D4LAPCHM, GMAOHIR1, GMAOHIR2, GMAOHIR3, GMAOHIR4
  - 272 GB/year
- MODIS
  - D4LAXMNT
  - 3 GB/year
- Andrew Klekociuk, Ray Wang
  - D4FAPMIS, D4LAPMIS; D4CAEMIS
  - 123 GB/year and 187 GB/year

Atmospheric Composition (Aura, SORCE)

Initial Project Review

Bruce Vollmer

Aura Evolution Agenda

- Requirements
  - Data Volume
- Interfaces
- Architecture
  - Software
  - Hardware
- Assumptions & Dependencies
- Test Plan & Schedule
- Risks
- Questions

Aura Requirements

OMI					
Processing Level	GB/day	TB/yr	Jan-06	Jan-07	Jan-08
0	6.6	2.4	3.6	6	8.4
1	16	5.8	8.7	14.5	20.3
2	3.1	1.1	1.7	2.8	3.9
2G	3.1	1.1	0	2.8	3.9
Total	28.8	10.4	14.0	26.1	36.5

Aura Requirements

MLS					
Processing Level	GB/day	TB/yr	Jan-06	Jan-07	Jan-08
0	1.2	0.5	0.7	1.2	1.7
1	4.0	1.5	2.2	3.7	5.2
2	0.8	0.3	0.5	0.8	1.1
3	0.1	0.04	0	0.1	0.15
Total	6.1	2.4	3.4	5.8	8.2

Aura Requirements

HIRDLS					
Processing Level	GB/day	TB/yr	Jan-06	Jan-07	Jan-08
0	0.8	0.3	0	0.6	0.9
1	1.0	0.4	0	0.8	1.2
2	0.1	0.04	0	0.08	0.12
3					
Total	1.9	0.75	0	1.5	2.3

Aura Requirements

SORCE					
Processing Level	MB/day	GB/yr	Jan-06	Jan-07	Jan-08
0	15	60	180	240	300
3	0.08	0.03	0.08	0.11	0.14
Total	15	60	180	240	300

Aura Requirements

- Ingest: FTP Pull, including secure interface where used
- Distribution Mechanisms: FTP Push/Pull, including secure interface where used
- Access Mechanisms: ECHO clients (WIST), WHOM, Mirador; restricted access to data yet to be validated
- Maintain Giovanni services for Aura online data
- Requirements for Archiving, Distribution and User Services in EOSDIS

Aura Interfaces

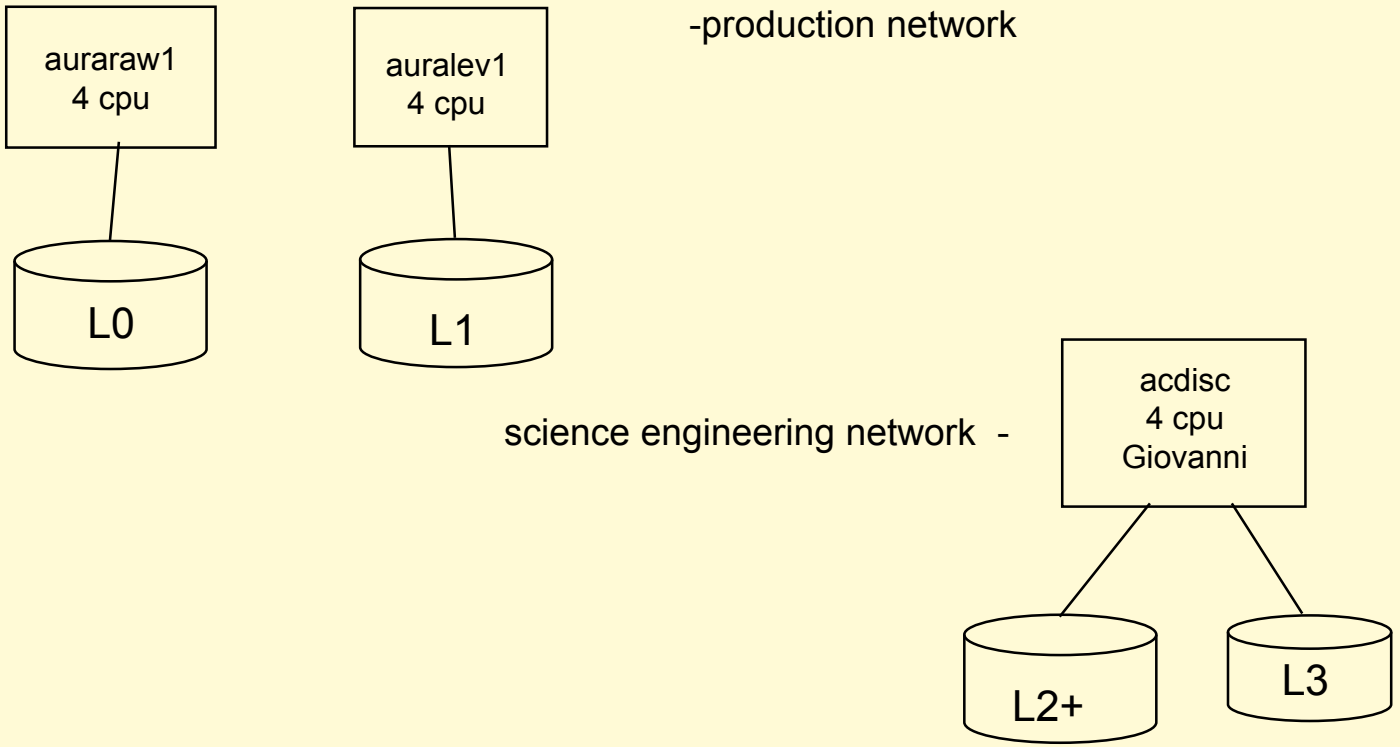
Interface	Data	Transfer Mechanism
EMOS	Input: att, eph, eng data	FTP Push
EDOS	Input: L0 science data	FTP Pull
NOAA, GMAO*	Input: dynamic ancillary data	FTP Pull
USNO	Input: UTCPoleT, LeapsecT	FTP Pull
OMI SIPS	Output: L0, DPREP, Input: L1, L2, L2G data	SCP Push* FTP Pull
ODPS	Output: L1, L2	SCP Push*
HIRDLS SIPS	Output: L0, DPREP, dynamic ancillary Input: L1, L2	FTP Pull FTP Pull
HIRDLS SCF	Output: L0, DPREP, UTCPoleT, LeapsecT	SCP Push*
BADC	Output: L0, DPREP, L1,L2, UTCPoleT, LeapsecT	FTP Push
MLS SIPS	Output: L0, DPREP, Input: L1, L2, L3 data	SCP Push* FTP Pull*
MLS SCF	Output: expedited data, dynamic ancillary data	FTP Push
SORCE	Input: L0, L3 data	SCP Push*
Users	Output: L2, L3 data	FTP Pull
*Secure interface		

Aura Architecture

- Software
  - S4PA software system with support for FTP Pull, FTP Push, secure shell ingest and distribution and access restrictions (where appropriate)
  - Giovanni running in Atmospheric Composition DISC

Aura Architecture (preliminary)

- Hardware



Aura Assumptions and Dependencies

- Two month turnaround for hardware procurement
  - PR submission to hardware delivery

Aura Test Approach

- Integrate, test, transition to Ops one team's set of interfaces at a time.
- This includes data from EDOS for a given team, data to and from a SIPS as well as associated facilities (SCFs)
  - OMI, ODPS, MLS, HIRDLS, SORCE

Aura Integration and Test Schedule

Start	Finish	
2/15/2006	3/31/2006	Aura S4PA Hardware design
5/1/2006	5/19/2006	Hardware procurement
5/23/2006	6/30/2006	Buy and ship hardware
7/15/2006	8/23/2006	Install and configure hardware
9/5/2006	9/22/2006	Integrate OMI S4PA interfaces (EDOS, SIPS, SCFs)
9/25/2006	10/6/2006	OMI S4PA end-to-end system testing
	10/10/2006	OMI S4PA interfaces Transition Readiness Review
10/11/2006	10/13/2006	OMI S4PA end-to-end test (mini-MOSS)
10/16/2006	11/3/2006	Integrate MLS S4PA interfaces (EDOS, SIPS, SCFs)
11/6/2006	11/21/2006	MLS S4PA End-to-end system testing
	11/27/2006	MLS S4PA interfaces Transition Readiness Review
11/28/2006	11/30/2006	MLS S4PA end-to-end test (mini-MOSS)

Aura Integration and Test Schedule

Start	Finish	
12/4/2006	12/21/2006	Integrate HIRDLS S4PA interfaces (EDOS, SIPS, SCFs)
1/2/2007	1/19/2007	HIRDLS S4PA end-to-end system testing
	1/22/2007	HIRDLS S4PA interfaces Transition Readiness Review
1/23/2007	1/25/2007	HIRDLS S4PA end-to-end test (mini-MOSS)
1/28/2007	2/9/2007	Integrate SORCE S4PA interface (LASP)
2/12/2007	2/23/2007	SORCE S4PA system testing
	2/26/2007	SORCE S4PA interface Transition Readiness Review
	2/27/2007	SORCE S4PA End-to-end test
	Note:	Schedule for Aura DPREP transition and GMAO data to Aura teams covered separately

Aura Risks

Likelihood	Consequence				
	Trivial	*Mild*	*Moderate*	*Critical*	*Catastrophic*
High (>0.5)					
Medium (0.1 -0.5)		**A**			
Low (0.01 -0.1)					
Unlikely (0.001 -0.01)					
Improbable (<0.001)					

A. Unknown reprocessing schedules	Aligning test and transition activities with reprocessing could impact schedules	Replan schedules as needed

Aura Risks

- No Medium, High risks identified at this time
- AIRS, GOLDS complete at start of Aura transitions
- Early success with OMI L2G in ACDISC

Aura Questions

- What are Aura teams reprocessing schedules?
  - ➢ Interface transition should align with reprocessing campaigns where possible, particularly SIPS to DISC data flows
    - Minimize potential user impacts
    - Minimize data migration activities
- What is the schedule for ODPS to DISC data transfers?

Terra & Ancillary Data Support Transition (TADS)

Initial Project Review

Gary T. Alcott

GES DAAC Operations Manager

TADS Agenda

- Requirements
- Interfaces
- Hardware Architecture
- Development Effort
- Test Plan
- Schedule
- Risk

TADS Requirements

- Terra Spacecraft
  - Receive and archive raw orbit and ephemeris data
  - Process and distribute orbit and ephemeris data
- Aqua/Aura Spacecraft*
  - Receive and archive raw orbit and ephemeris data
  - Process and distribute orbit and ephemeris data
- ASTER Expedited
  - Receive, store, and distribute ASTER Expedited Data
  - ASTER Email Gateway
- Ancillary Products
  - Receive, archive, and distribute ancillary products
  - Process, archive, and distribute select products into HDF format

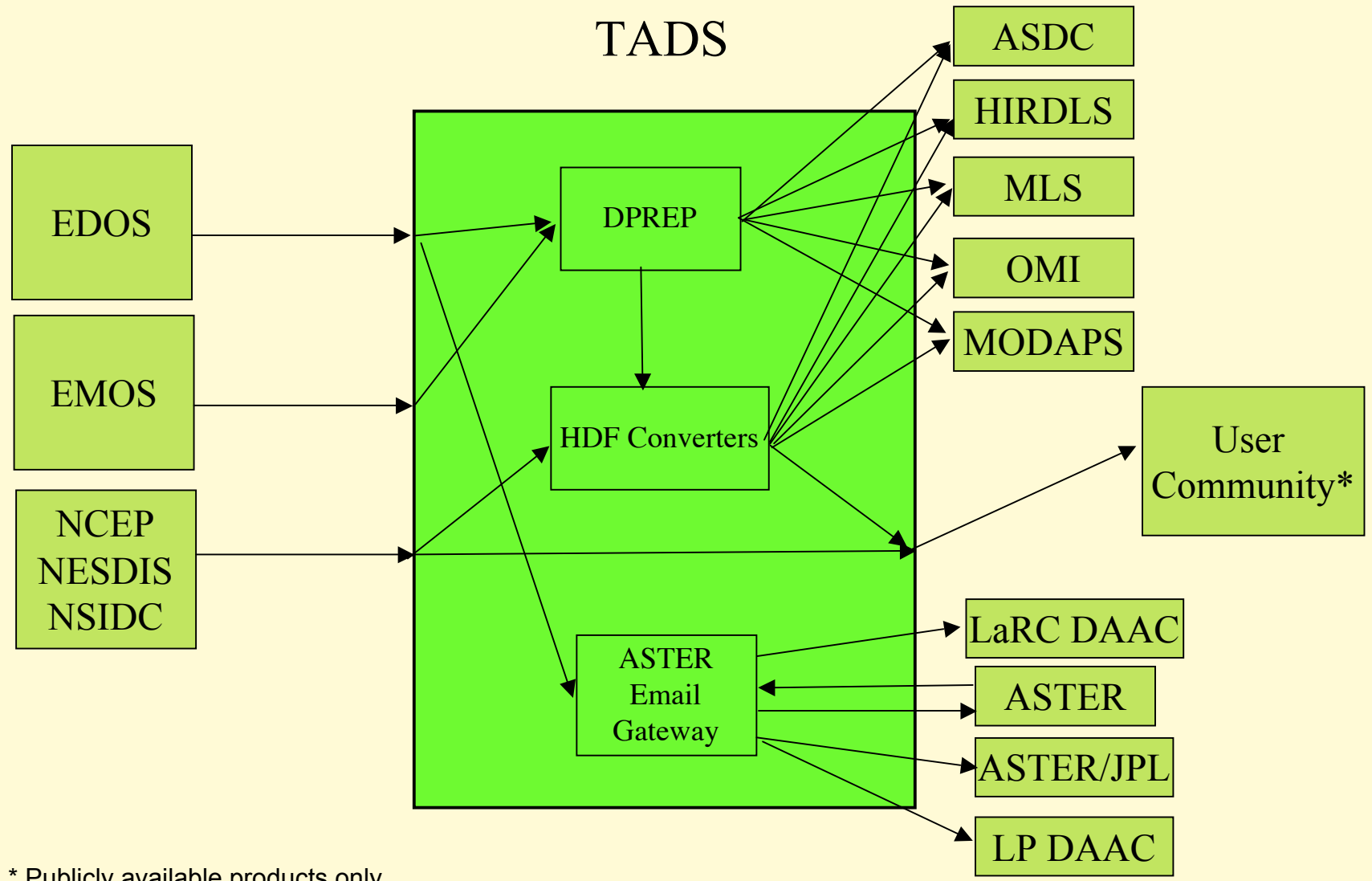
* Aqua DPREP processing, archive, and distribution will take place on the AIRS instance of S4PA ,but utilizes the same code as the Aura instance and therefore will not require separate development.

TADS Data Volumes

Data Group	GB/day	TB/year	Jan-06	Jan-07	Jan-08
Terra Spacecaft	0.23	0.08	0.44	0.52	0.60
Aura Spacecraft	0.10	0.04	0.07	0.10	0.14
ASTER Expedited	0.15	0.07*	0.07*	0.07*	0.07*
Ancillary (Native)	0.40	0.14	0.84	0.98	1.13
Ancillary (HDF)	1.23	0.43	2.27	2.71	3.14
Total	2.11	0.69*	3.69	4.38	5.08

* ASTER Expedited data is kept a rolling 45-day storage

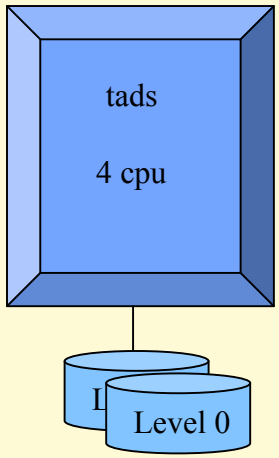
TADS Interfaces



* Publicly available products only

TADS Architecture

Terra & Ancillary Data Support (TADS)



Software:
- S4PA
- DPREP
- ASTER Email Gateway
- HDF Converters

TADS Development Effort

- Port of DPREP code to S4PA
  - Terra DPREP of orbit and ephemeris data
  - Aqua/Aura DPREP of orbit and ephemeris data
- Port of HDF Converters
  - For converting DPREP output and 12 ancillary data types
  - Turn into a PGE
- ASTER Email Gateway
  - Develop scripts to emulate existing interface
    - Send notices for received expedited data sets (EDSs)
    - Receive ASTER requests for selected EDSs
    - Delivery ASTER EDSs
    - Subscriptions of all EDSs to LP DAAC and ASTER JPL

TADS Test Plan

- Local generation and verification of output for each product type in comparison to existing processing.
- Verification of sample products of each product type by end users.
- Interface testing of product flow from TADS with end users in parallel with current flow from ECS
- Run Mini-MOSS followed by TRR
- Switch operational flow to TADS with ECS poised as a backup
- Turn off ECS functionality

TADS Schedule

Start*	Finish	
11/27/06	01/05/07	HDF Converters Port
05/07/07	05/11/07	HDF Converters Mini-MOSS/TRR
04/02/07	05/24/07	HDF Converters Transition
01/08/07	01/26/07	Aqua/Aura DPREP Port
06/01/07	06/07/07	Aura DPREP Mini-MOSS/TRR
04/13/07	06/21/07	Aura DPREP Transition
02/09/07	03/22/07	Terra DPREP Port
05/31/07	06/06/07	Terra DPREP Mini-MOSS/TRR
04/26/07	06/20/07	Terra DPREP Transition
03/23/07	05/03/07	ASTER Email Gateway Development
06/18/07	06/22/07	ASTER EDS Mini-MOSS/TRR
05/14/07	07/20/07	ASTER EDS Transition
04/02/07	05/11/07	Provider Interface Testing

* Some items will start earlier as risk mitigation.

TADS Risks

Likelihood	Consequence				
	Trivial	*Mild*	*Moderate*	*Critical*	*Catastrophic*
High (>0.5)					
Medium (0.1-0.5)					
Low (0.01-0.1)				**A**	
Unlikely (0.001-0.01)					
Improbable (<0.001)					

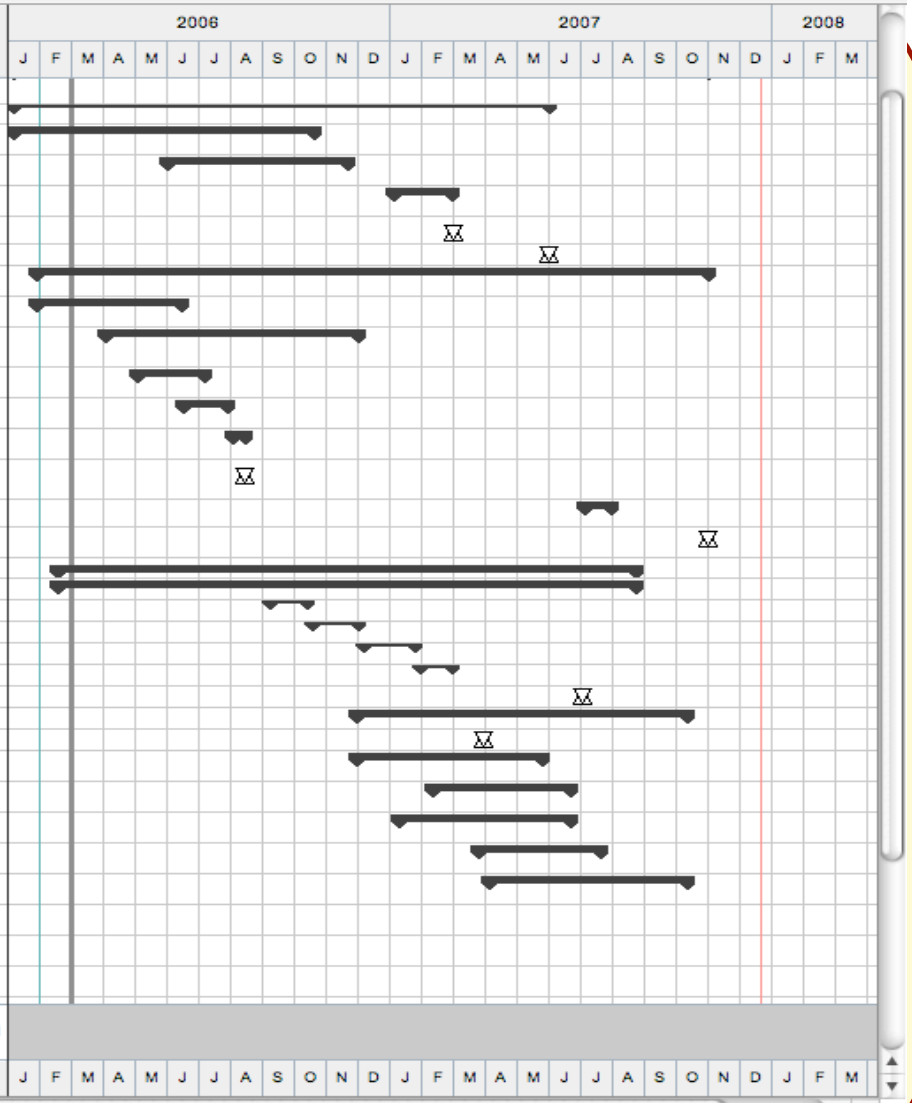
A. Port of DPREP, PREPQC, and HDF Converters to S4PA	Encountered dependency on RogueWave Libraries. Other dependencies may exist.	*Research: Successfully created a local DPREP development environment and compiled all DPREP code.* *Mitigate: DPREP developer is part of DAAC engineering staff.* *Mitigate: ECS developer to provide RogueWave source code and to provide RogueWave port to Linux*

Wrap-Up

Wrap-Up

GES DISC Miles3/2.fts

	WBS	Activity Name	Duration (Work Days)	Start Date	Finish Date
10					
11	2.1	**Development Schedule: Data AIRS L0 Data Switchover to V3**	368.00	1/4/06	6/1/07
12	2.1.1	*Preparation...*	208.00	1/4/06	10/20/06
28	2.1.2	*Implementation...*	125.00	5/30/06	11/20/06
47	2.1.3	*Transition live interfaces...*	41.00	1/2/07	2/27/07
54	2.1.4	*AIRS Transition Readiness Review*	0.00	2/28/07	2/28/07
55	2.1.5	***Turn off ECS Feed***	1.00	6/1/07	6/1/07
56	2.2	**GMAO Mission Migration - G4/G5**	459.00	1/27/06	11/1/07
57	2.2.1	*Hardware...*	98.00	1/27/06	6/13/06
62	2.2.2	*Need by dates for S4PA S/W development...*	175.00	3/31/06	12/1/06
70	2.2.3	*Subsetter porting...*	48.00	5/1/06	7/5/06
74	2.2.4	*GOLDS Flows...*	33.00	6/14/06	7/28/06
79	2.2.5	*GEOS-5 Mini-MOSS...*	10.00	8/1/06	8/14/06
84	2.2.6	*ECS not needed for GEOS-4 (except for CERES)*	0.00	8/15/06	8/15/06
85	2.2.7	*GEOS-4 Flow to GOLDS...*	20.00	7/2/07	7/30/07
93	2.2.8	***ECS not needed***	0.00	11/1/07	11/1/07
94	2.3	**Aura L0 Data Switchover**	397.00	2/15/06	8/23/07
95	2.3.1	Hardware...	397.00	2/15/06	8/23/07
100	2.3.2	*OMI...*	29.00	9/5/06	10/13/06
105	2.3.3	*MLS...*	34.00	10/16/06	11/30/06
110	2.3.4	*HIRDLS...*	39.00	12/4/06	1/25/07
115	2.3.5	*SORCE...*	22.00	1/29/07	2/27/07
120	2.3.6	***Turn off ECS Feed***	1.00	7/2/07	7/2/07
121	2.4	**TADS**	229.00	11/27/06	10/11/07
122	2.4.1	*TADS Hardware Available*	1.00	3/30/07	3/30/07
123	2.4.2	*HDF Converters...*	129.00	11/27/06	5/24/07
132	2.4.3	*Terra Orbit & Ephemeris...*	94.00	2/9/07	6/20/07
141	2.4.4	*Aqua/Aura Orbit & Ephemeris...*	119.00	1/8/07	6/21/07
151	2.4.5	*ASTER EMail Gateway...*	86.00	3/23/07	7/20/07
159	2.4.6	*TADS Interface Testing...*	139.00	4/2/07	10/11/07



Wrap-Up

- Software and hardware development underway
- Next Steps
  - Continue working with ITs to ensure evolution of interfaces is seamless and non-intrusive to IT
  - Procure 2006 hardware
  - Transition MODIS data management

Wrap-Up

- Status:
  - Work breakdown is in place
  - Schedules are in place
  - Personnel have/will be assigned
  - Communications with stakeholders are ongoing
  - Plans and procedures (new and reused) are in place
  - Risks have been identified

AIRS Interfaces

Interface	Data	Transfer Mechanism
EMOS EMOS (from FDD)	Input: attitude, engineering data Input: ephemeris	FTP Pull FTP Push
EDOS	Input: L0 ancillary, science data	FTP Push
NOAA	Input: dynamic ancillary data	FTP Pull
USNO	Input: UTCPoleT, LeapsecT data	FTP Pull
AIRS SCF	Input: Algorithm Software Output: Selected L0, L1-L3	FTP Push
NSIDC	Output: Aqua DPREP output (AMSR-E)	FTP Pull
Users	Output: L1-L3, Browse	FTP Push, Pull

Aura Interfaces

Interface	Data	Transfer Mechanism
EMOS EMOS (from FDD)	Input: attitude, engineering data Input: ephemeris	FTP Pull FTP Push
EDOS	Input: L0 science data	FTP Push
NOAA, GMAO*	Input: dynamic ancillary data	FTP Pull
USNO	Input: UTCPoleT, LeapsecT	FTP Pull
OMI SIPS	Output: L0, DPREP, Input: L1, L2, L2G data	SCP Push* FTP Pull
ODPS	Output: L1, L2	SCP Push*
HIRDLS SIPS	Output: L0, DPREP, dynamic ancillary Input: L1, L2	FTP Pull FTP Pull
HIRDLS SCF	Output: L0, DPREP, UTCPoleT, LeapsecT	SCP Push*
BADC	Output: L0, DPREP, L1,L2, UTCPoleT, LeapsecT	FTP Push
MLS SIPS	Output: L0, DPREP, Input: L1, L2, L3 data	SCP Push* FTP Pull*
MLS SCF	Output: expedited data, dynamic ancillary data	FTP Push
SORCE	Input: L0, L3 data	SCP Push*
Users	Output: L2, L3 data	FTP Pull
*Secure interface		