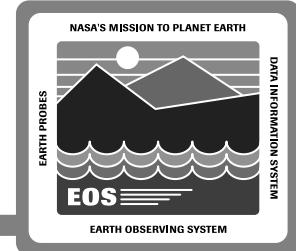


Ingest Subsystem

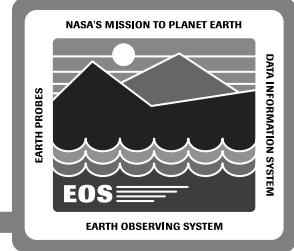
C. Gire

15 February 1995

Ingest Subsystem Agenda



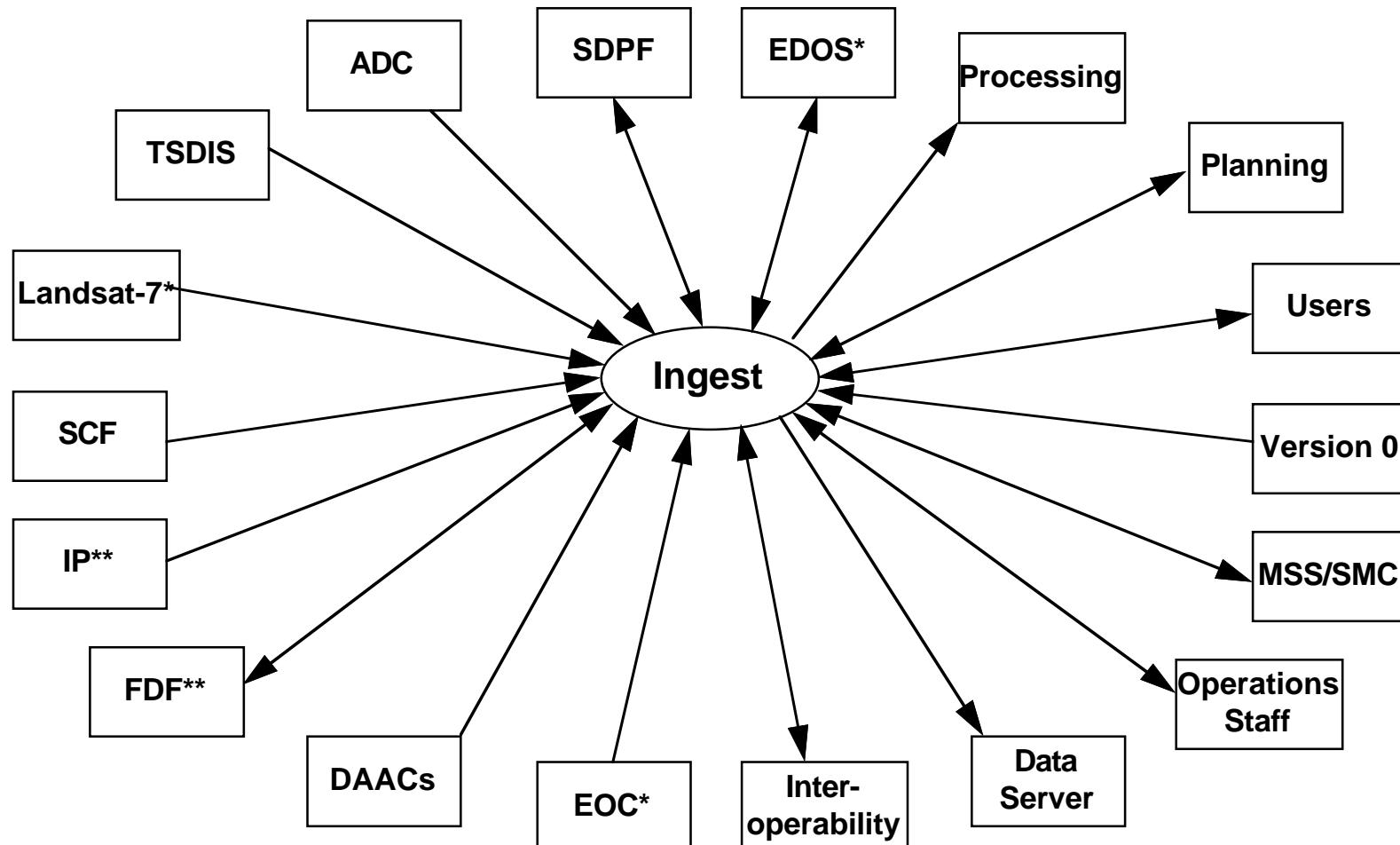
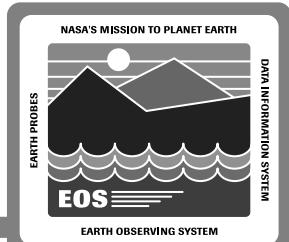
- Overview
- Design Drivers
- Design Approach/Rationale
- Design Concepts/Issues
- Data Ingest Scenario
- Ingest Subsystem Software
- Ingest Subsystem Hardware
- Phasing of Capabilities
- Follow-on Engineering Activities



Ingest Subsystem Overview

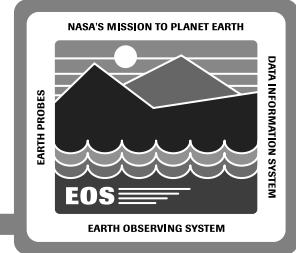
- Responsible for ingest of data from external data providers (SDPF, TSDIS, EDOS, Landsat-7, etc.)
 - Data transfer and transmission checking
 - Data conversions (as required)
 - Metadata extraction (as required)
 - Critical metadata checking
 - Storage of Level 0 data for one year
 - Provides Level 0 data to Data Processing
 - Inserts higher-level data into the Data Server
- Provides single point of operations staff monitoring and control of data ingested from external data providers
- Composed of the Ingest CSCI (INGST) and the Ingest Client HWCI (ICLHW)

Ingest Subsystem Context Diagram

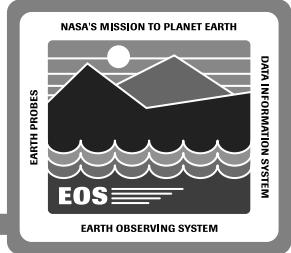


Note: * = early interface test after Release A; ** =

Ingest Subsystem Design Drivers



- Interfaces:
 - Very high reliability ingest of Level 0 (SDPF and EDOS)
 - High-volume Level 0 data ingest; 1-year Level 0 data storage
 - Wide variety of other interfaces (TSDIS, NOAA, Landsat-7, etc.)
 - Addition of new interfaces and evolvability of existing interfaces
 - Version 0 data migration
- Operations:
 - Single point of operations monitoring and control
 - Tunable ingest parameters to support ingest load balancing
 - Maximal cost-effective use of automation
- Exception handling
 - Standard LSM interface to error log with alerts (e.g., flashing icon) for critical faults
 - Component-level recovery after failures

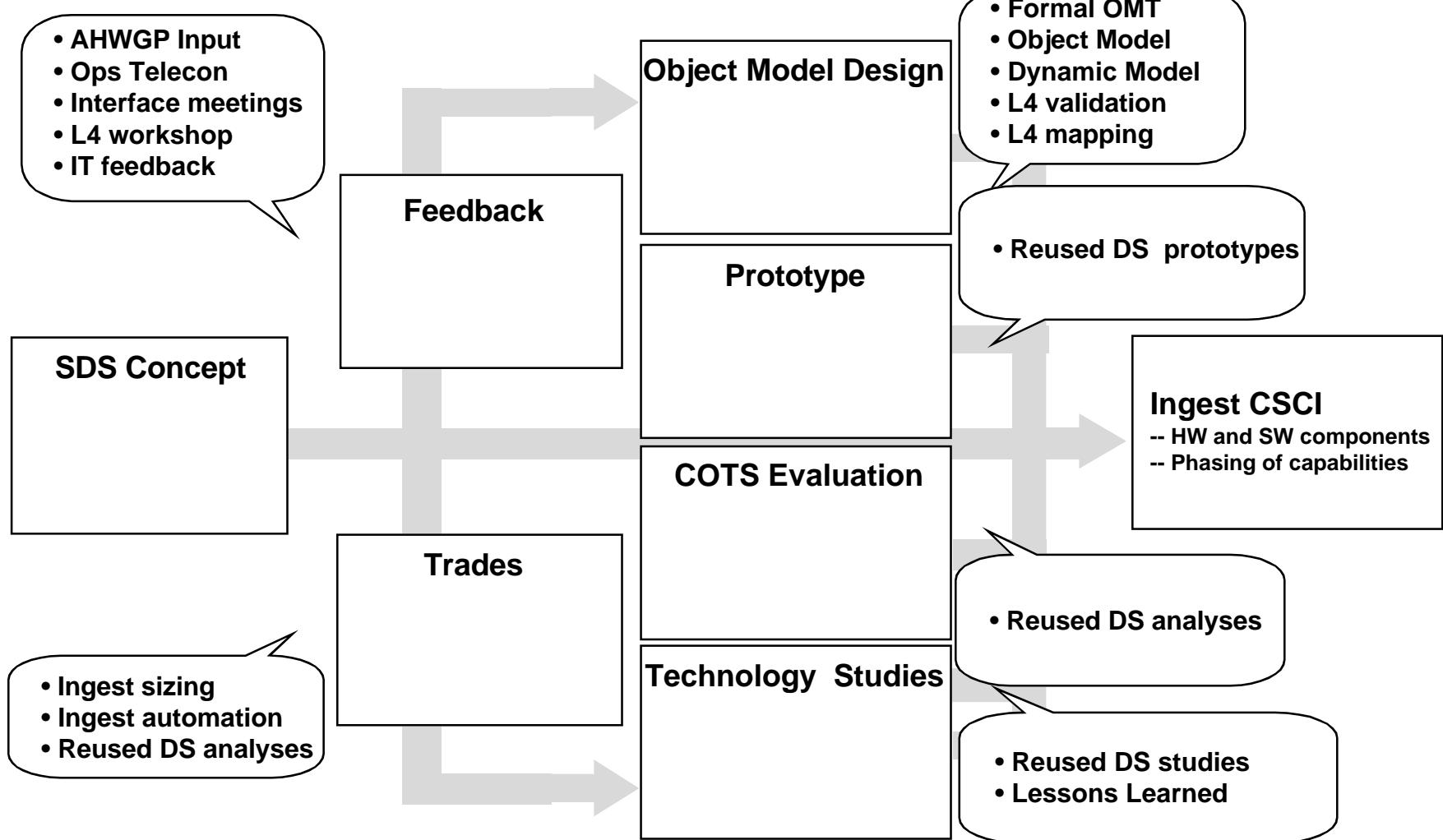
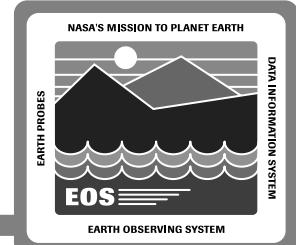


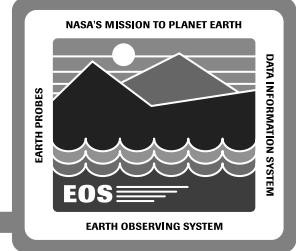
Ingest Data Size Drivers

| Interface | Volume | Frequency |
|-------------------------------|---------------|-----------|
| SDPF - CERES L0 | 90 MB | Per day |
| SDPF - LIS L0 | 65 MB | Per day |
| TSDIS - All data | 13.7 GB | Per day |
| DAO - NMC data | 22 MB | Per day |
| NESDIS - All data | 64 MB | Per week |
| NOAA (GPCP, GPI) | 5.1 MB | Per month |
| GSFC DAAC (TOMS) | TBD | TBD |
| LaRC DAAC (SAGE-II, ISSCP) | TBD | TBD |
| MSFC DAAC (SSM/I) | 168 MB | Per day |
| EDOS - GSFC (AM-1) | 67 GB | Per day |
| EDOS - LaRC (AM-1) | 41.4 GB | Per day |
| EDOS - EDC (ASTER) | 89 GB (media) | Per day |
| Landsat-7 - L0R | 140 GB | Per day |

Note: Only interfaces with large volumes are shown for non-Release A interfaces

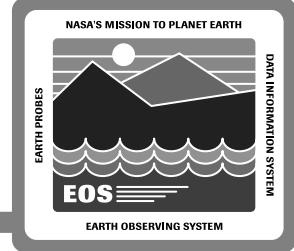
Ingest Design Approach





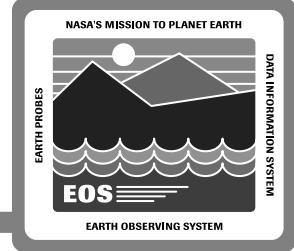
Ingest Design Rationale

- Design Analyses
 - Ingest Subsystem sizing study
 - “Man-in-the-loop” ingest automation study
- Design Trades
 - Tie-in to Data Server trade studies (FSMS, working storage, data repository technology)
- Prototypes
 - Tie-in to Data Server prototypes (FSMS, data repository technology)
- COTS Evaluations
 - Tie-in to Data Server DBMS evaluations



Key Ingest Design Concepts

- Separate, highly-reliable hardware components for Level 0 data ingest
 - Ingest “rolling storage” of Level 0 data
- Reuse of Data Server software to support Level 0 data storage and retrieval
- Standardization of external interface hand-shaking
 - Automated, data availability notice (DAN)-driven (SDPF, TSDIS, DAACs, Landsat-7, SCF, EOC, CSMS, future interfaces)
 - Automated, polling interval-driven (EDOS, DAO, NESDIS)
 - Media-driven (ASTER Level 1 data, Version 0 data, backup data)
 - Authorized science user interactive network ingest (Version 0 data)
- Managed list of ingest requests; single point of monitoring and control

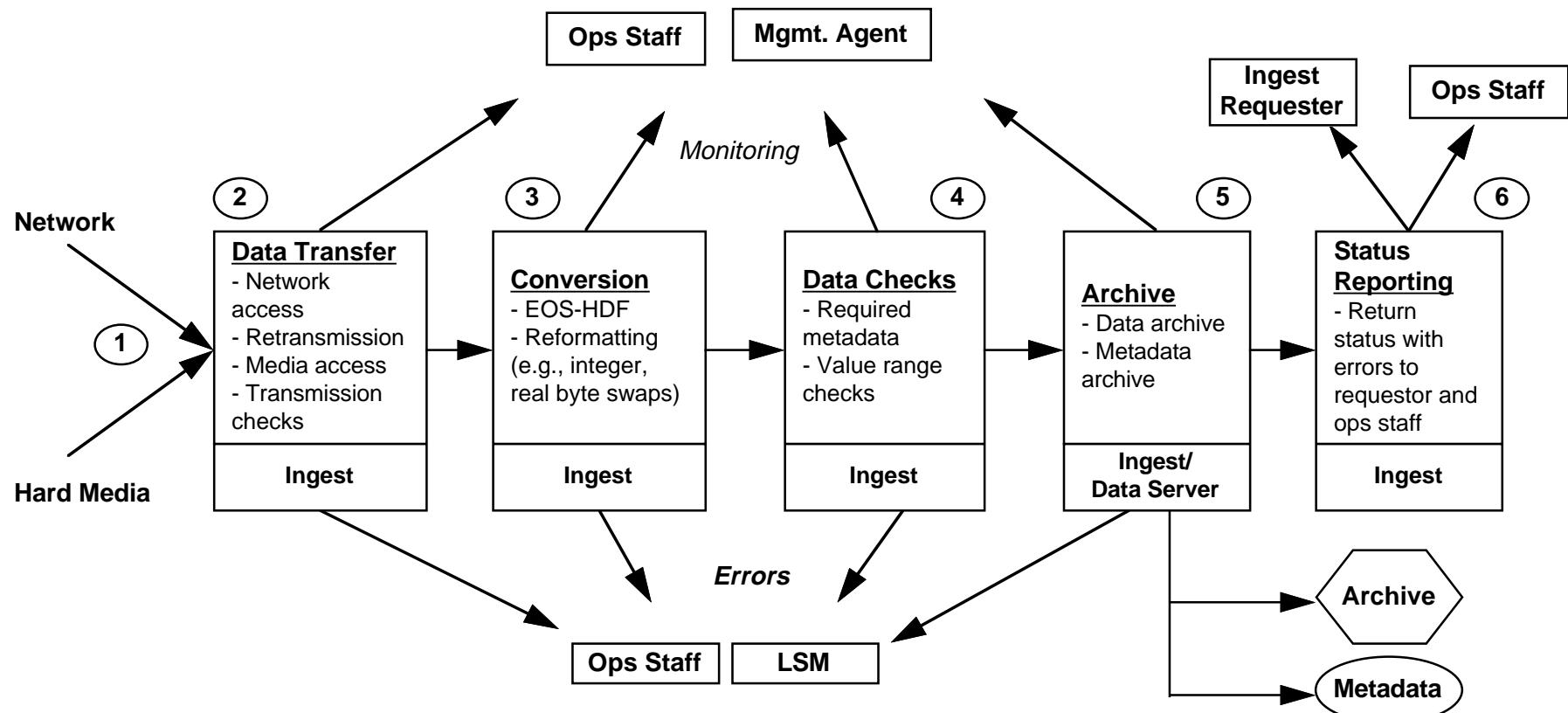
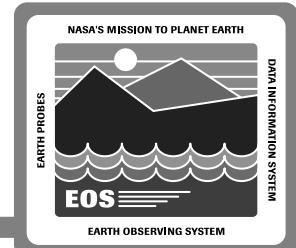


Key Ingest Interface Issues

- Potential for additional interface standardization
- Standard Delivery Record file format for polling ingest and media ingest
- Final Release A interface definitions (e.g., data volumes, formats) at CDR
- EDOS buffer size/Production Data Set (PDS) definition
- Landsat-7 interface ingest window (<N hours per pass)
- Flight Dynamics Facility (FDF) interface

A

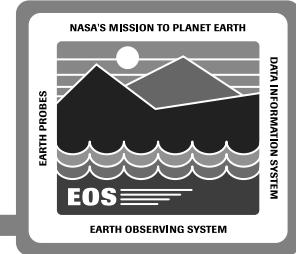
Data Ingest Scenario



Note: this is a general scenario for ingest of Level 0 and higher-level data

Representative Data Ingest Steps

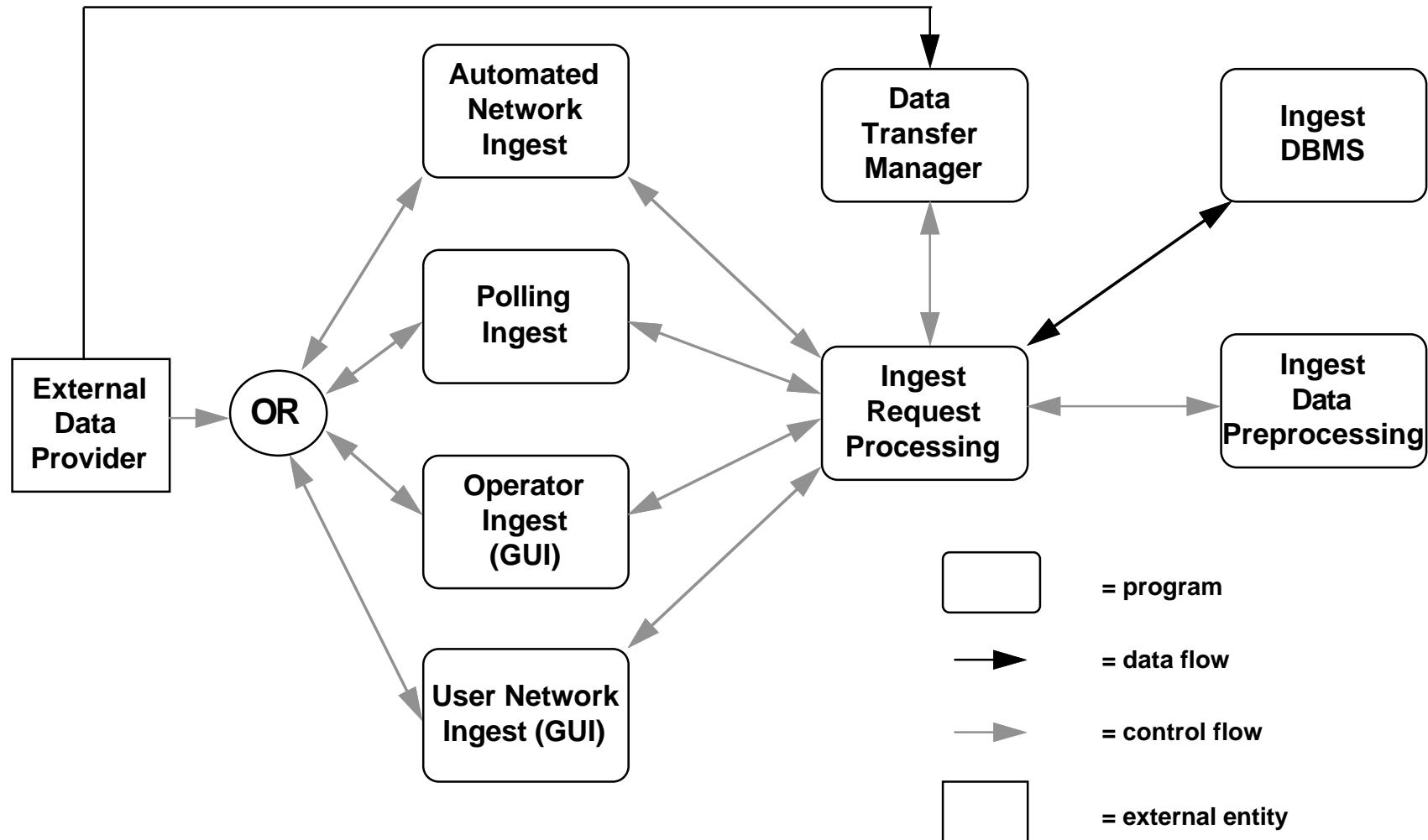
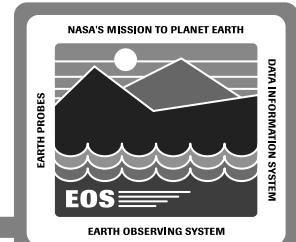
SDPF Level 0 Data Ingest



- 1) SDPF sends data availability notice to ECS**
 - Ingest CSCI client generates a corresponding ingest request
 - Ingest CSCI stores the request on a prioritized list and determines the estimated time required to ingest the data
 - Ingest CSCI client acknowledges the request
 - Operations staff may monitor subsequent ingest request processing and cancel (/suspend/resume A) requests
- 2) Ingest CSCI schedules data ingest and performs data transfer**
- 3) No data conversion is required for SDPF Level 0 data**
- 4) Ingest CSCI extracts SDPF Level 0 metadata and checks selected metadata parameters (e.g., SFDU header information)**
- 5) Ingest CSCI requests to store SDPF Level 0 data and metadata in the Level 0 data repository**
- 6) Ingest CSCI client reports status to the requester and to the Ingest History Log when the data storage is complete**

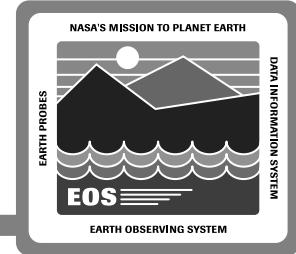
Ingest Subsystem

Ingest CSCI Component Diagram



Ingest Subsystem

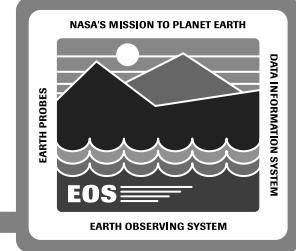
Ingest CSCI



- Automated Network Ingest Client
 - Provides hand-shaking for DAN-driven automated ingest
- Polling Ingest Client
 - Provides polling interval-driven automated ingest
- Ingest Request Processing Manager
 - Manages ingest request list
- Data Transfer Manager (aka Ingest Working Collection)
 - Manages network and media data transfers
- Ingest Data Preprocessing
 - Manages ingest data preprocessing (conversions, reformatting, metadata extraction and checking, Data Server interface)
- Operator Ingest (GUI)
 - Provides GUI monitoring and control of ingest processing
 - Media ingest interface Client

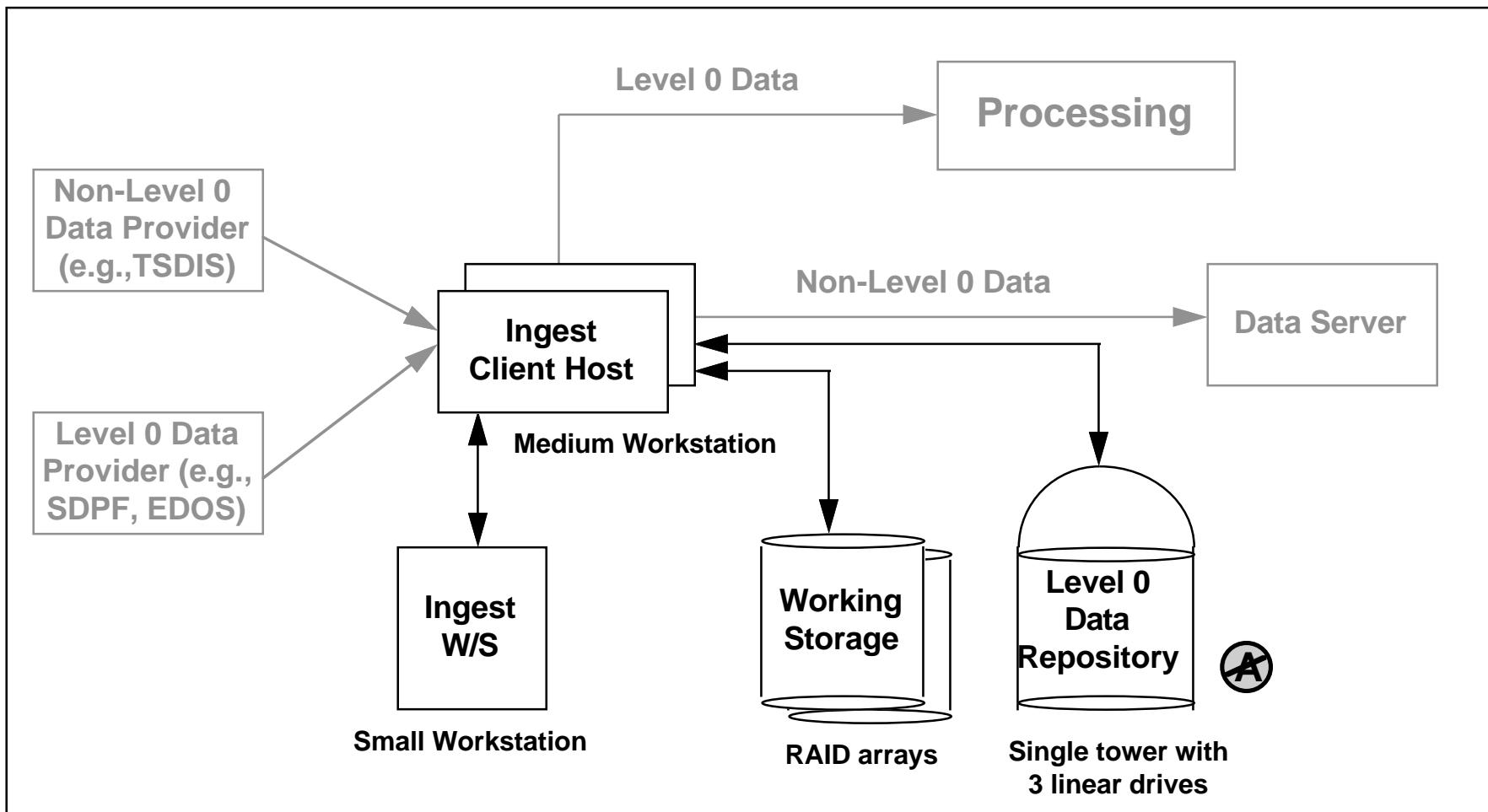
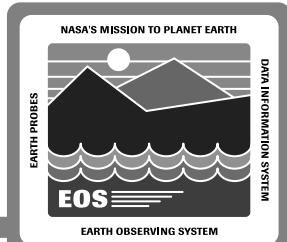
Ingest Subsystem

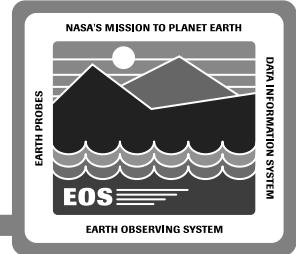
Ingest CSCI (cont.)



- User Network Ingest (GUI) Client
 - Provides GUI interactive ingest and status monitoring
- Ingest DBMS
 - Provides data base storage of ingest tables
- Ingest File Storage Software
 - Manages Level 0 data repository--reuse from Data Server
- Resource Administration Application
 - Provides GUI monitoring and control of Level 0 data repository--reuse from Data Server
- Media Handling Software
 - Provides media (e.g., 8mm tape) access--reuse from Data Server
- Viewing Tools
 - Provides visualization and analysis tools--reuse from Client

Ingest Client HWCI (ICLHW) Hardware Diagram--Release A

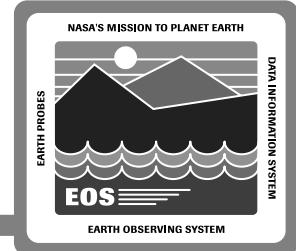




Ingest Client HWCI

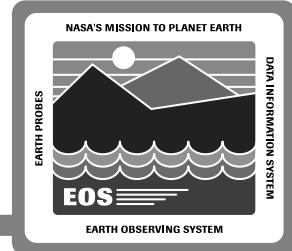
- Ingest client hosts
 - High-reliability, availability
 - Network interfaces for data ingest
 - Sized to support required ingest data preprocessing
- Working storage
 - Sized to support high-availability Level 0 data ingest
- Level 0 data repository
 - Provides Level 0 “rolling storage” (storage of Level 0 data for 1 year)
 - Level 0 repository implemented with working storage devices at smaller sites
- Ingest workstations
 - Support monitoring and control of external data ingest
- All critical ingest components are cross-strapped in warm spare or redundant configurations to satisfy RMA requirements

Ingest Client HWCI Scalability



- Ingest client hosts
 - Upgrade within existing processor class; upgrade to higher processor class; add new processor
 - Dependent on I/O estimates and ingest data preprocessing
- Working storage
 - Addition of RAID arrays/larger capacity arrays
- Level 0 data repository
 - Addition of tape recorders (for additional I/O throughput)
 - Addition of robotics units (for additional data volume)
- Ingest workstations
 - Additional workstations

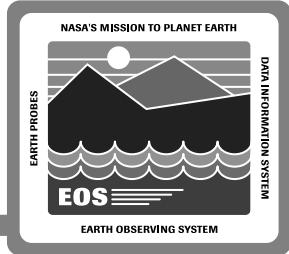
Ingest Subsystem CSCI-to-HWCI Mapping



| CSCI Component | HWCI Component* |
|------------------------------|-------------------|
| Automated Network Ingest | ICLHW host |
| Polling Ingest | ICLHW host |
| Operator Ingest | ICLHW workstation |
| User Network Ingest | User Workstation |
| Ingest Request Processing | ICLHW host |
| Ingest Working Collection | ICLHW host |
| Ingest Data Preprocessing | ICLHW host |
| Ingest DBMS | ICLHW host |
| Ingest File Storage Software | ICLHW host |
| Media Handling Software | ICLHW host |
| Viewing Tools | ICLHW workstation |
| Resource Admin. Application | ICLHW workstation |

*Note: Dependent on DAAC sizing requirements, some or all h/w components may merge with Data Server components.

Ingest Subsystem Phasing of Capabilities



Interim Release 1 (IR-1):

- Automated Network Ingest--SDPF and TSDIS interface testing
- Polling Ingest--NESDIS and DAO interface testing

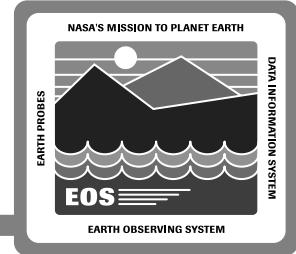
Release A:

- Ingest Request Processing
- Ingest Working Collection
- Ingest Data Preprocessing (metadata extraction/checking)
- Operator Ingest GUI (media ingest, request monitoring, cancellation)
- User Network Ingest GUI (interactive network ingest, monitoring)

Release B:

- Ingest Data Preprocessing (data conversions, reformatting)
- Operator Ingest GUI (request suspension, resumption)

Ingest Subsystem Follow-on Engineering Activities



Development:

- IR-1 development (review detailed design in May)
- Release A development (CDR in August)
- Release B preliminary design (IDR in September)

Trade Studies/Analyses/System Engineering:

- Data conversion/reformatting analysis
- Final data volume/frequency modeling
- IR-1 and Release A hardware specification and procurement
- Finalization of external ICDs

Prototypes:

- Level 0 rolling storage prototype (in conjunction with Data Server)
- Level 0 ingest failover prototype
- Operator GUI prototype
- SDPS infrastructure prototype