

---

# Ingest Subsystem Overview

**Carey Gire**

**[cgire@eos.hitc.com](mailto:cgire@eos.hitc.com)**

---

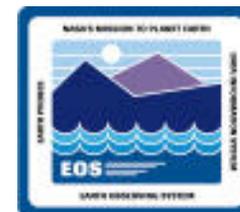
**15 April 1996**



# Presentation Outline

- **Driving Requirements**
- **External Interfaces**
- **ECS Context/Release B Capabilities**
- **IDR & Workshop Issue Work-off**
- **Prototypes and Trades**
- **COTS Selections**
- **Hardware/Software Architecture**
- **Software Design**
- **Data Base Design**
- **Ingest Recovery**
- **OMT Review -- Object Models, Event Traces, PDL**
- **305 Errata**

# Ingest Subsystem Driving Requirements



## Functional Requirements

- Receive data from external data providers
- Extract metadata, convert data as required, provide to Data Server
- Provide operator monitoring and control of ingest processing

## Reliability, Maintainability, Availability (RMA)

- 0.999 availability, 20 minute mean down time (MDT)
- 15 minute switchover to backup capability
- Storage of L0 data

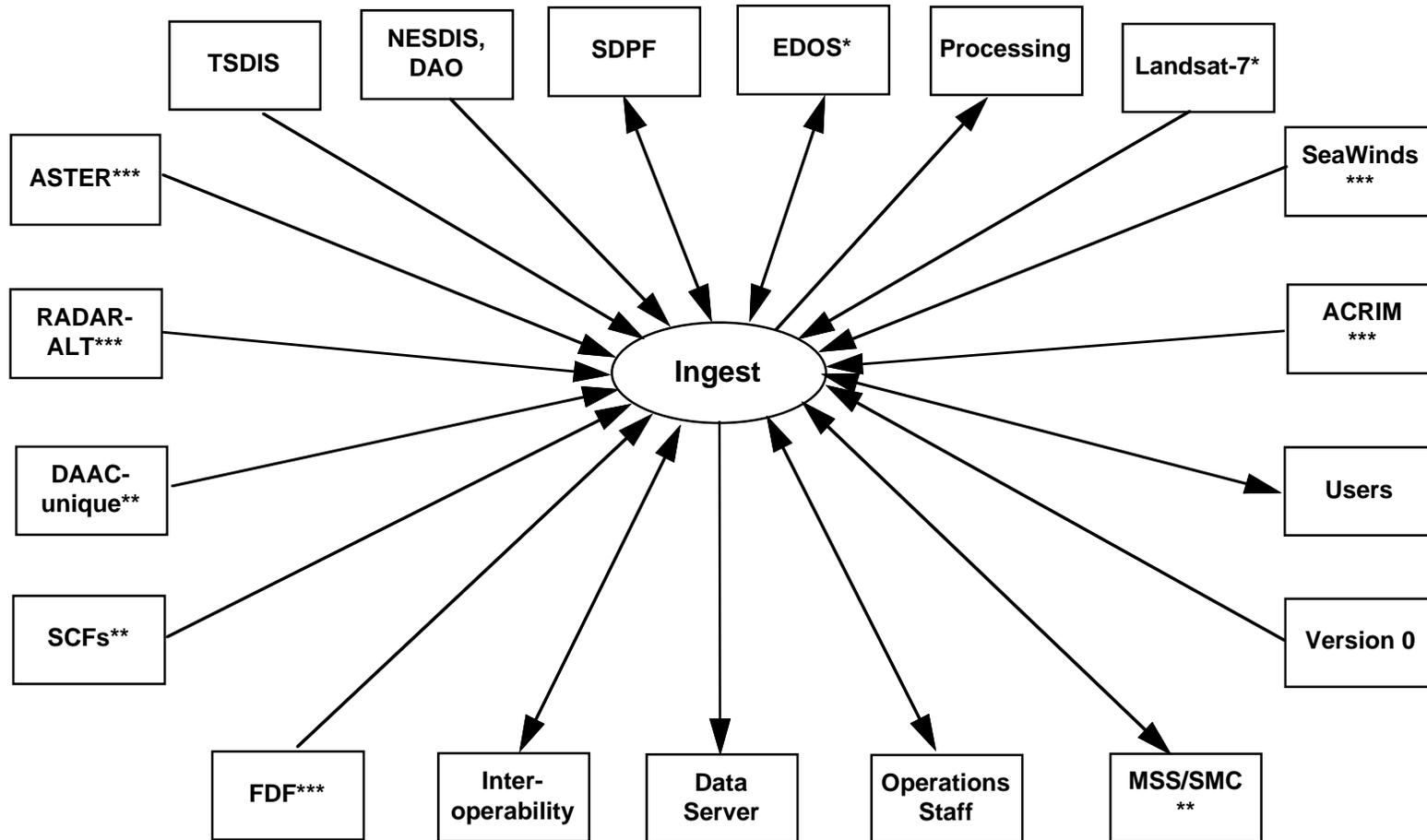
## Data Integrity

- Authorized access by data providers
- Data transmission checking
- Metadata checking

## Scalability, Evolvability

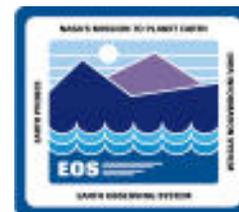
- Large, increasing number of interfaces, data volumes, I/O rates
- Capability to handle 3X daily I/O load
- Ongoing Version 0 (V0) data migration

# Ingest Subsystem Context Diagram



Note: \* = early interface test using Release A; \*\* = augmented at Release B; \*\*\* = Release B

# Ingest Interface Issues/ Resolutions



## **EDOS Interface**

- Interface issues are converging
- EGHS header overlaid onto PDS Delivery Record structure--no immediate impact to ECS
- D3 media for L0 recovery recently proposed; design material update required
- Ongoing meetings with EDOS engineers; ICD need date--6/3/96

## **Meteor/SAGE-III Interface**

- Preliminary interface descriptions in place
- Final ICD expected in June, 1996; assume minimal impact due to use of existing interface approaches, formats

## **Landsat-7 IGS Interface**

- Ingest media type not finalized (expect 8mm tape); need date
- Metadata and browse formats not defined (expect PVL and ECS standard browse format)
- ICD update required; need date 8/1/96

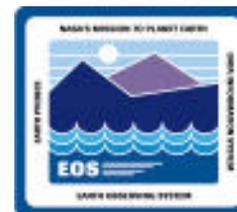
## **ASTER Interface**

- D3 media for L1a/L1b delivery recently finalized; design material update required

## **NESDIS (NOAA SAA) Interface**

- Polling without delivery record approach is under review; issue for Release A that is converging

# Ingest Interface Issues/ Resolutions (cont.)



## ORNL Interfaces

- No metadata defined; need date 8/1/96

## ASF Interfaces

- Working version of ICD; issues are being worked in support of “bridge analysis plan”

## SeaWinds Interface

- Preliminary IRD in place
- Late delivery of ICD expected in December, 1996; requires replan of system test; assume minimal impact due to use of existing interface approaches, formats

## RADAR ALT, ACRIM Interfaces

- Still undefined; need date 8/1/96; assume minimal impact due to use of existing interface approaches, formats

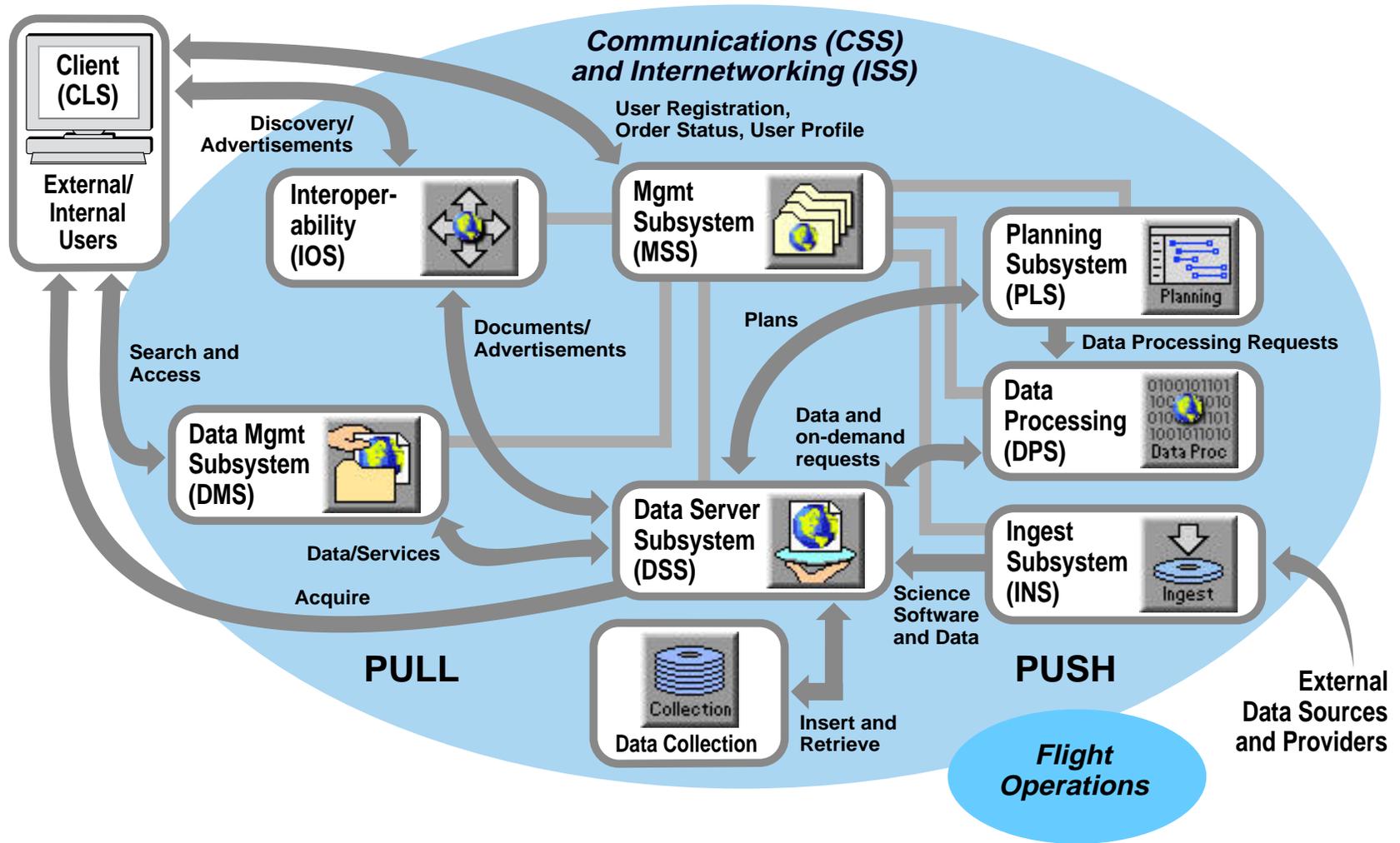
## FDF Interface

- ICD not yet available from FDF; ingest interface limited to potential media delivery on a very infrequent basis for repaired ephemeris data

## User Ingest Interfaces

- Refining process for defining new authorized data types, characteristics (e.g., via an interface definition process similar to that used for institutions); need date 8/1/96

# ECS Context





# Key Ingest Design Concepts

**Ingest is composed of the ICLHW HWCI and the INGST CSCI**

**Separate, highly-reliable hardware components for Level 0 data ingest**

- **Ingest archive repository for Level 0 data**

**The INGST CSCI is entirely on the formal track; small amount of custom software for Release B**

**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, Landsat-7, SCF, MSS)**
- **Automated, polling interval-driven (EDOS, ACRIM, RADAR ALT, SAGE III, ADEOS II, DAACs, plus NESDIS)**
- **Media-driven (ASTER, Version 0 data, backup data)**
- **Authorized science user interactive network ingest (SCF, Version 0)**

**Automated retries, semi-automated component failover**

**Managed list of ingest requests; single point of monitoring and control**



# Release B Ingest Capabilities

## External Interfaces:

- Existing IFs with TSDIS, SDPF, DAO, NESDIS, LaRC and GSFC DAAC, and CERES SCF
- New interfaces with EDOS, Landsat 7 LPS, Landsat 7 IAS and IGSs, ASTER, ADEOS II, RADAR-ALT, SAGE III, ACRIM, additional SCFs, and other DAAC-unique IFs

## Ingest Control:

- Release A capability to submit requests, cancel requests, checkpoint requests, recover requests
- Release B capability to suspend/resume requests or classes of requests (by external data provider, or all requests), change request priority

## Ingest Performance Requirements:

- Release A data provider volumes and I/O rates (< 20 GB per day)
- Additional Release B data provider volumes and I/O rates (up to 140 GB per day)

# Release B Ingest Capabilities (cont.)



## Data Preprocessing:

- **Metadata extraction for Release A and Release B data types**
- **Conversions of NMC ancillary data**
- **Additional Release B ancillary data conversions still under evaluation**
  - Need date: 6/3/96
  - See Appendix A for a list of known ancillary data

## Miscellaneous:

- **HTML forms for Release A guide data, science software archive package**
- **HTML forms for additional Release B data types**
- **Additional media ingest capabilities (ASTER)**
- **Hardcopy digitizing/scanning**



# Ingest Data Size Drivers

Interface	Release A Volume Per Day	Release B Volume Per Day
<b>GSFC - Release A Sources (incl. TSDIS)</b> <b>GSFC - EDOS (MODIS L0)</b> <b>GSFC - ACRIM L0/others</b> <b>GSFC - V0 Migration</b> <b>LaRC - Release A Sources (incl. TRMM CERES L0)</b> <b>LaRC - EDOS (CERES L0, MISR L0, MOPITT L0)</b> <b>LaRC - V0 Migration</b> <b>EDC - Landsat-7 L0R</b> <b>EDC - ASTER - L1A/L1B (media)</b> <b>EDC - V0 Migration</b> <b>JPL - RADAR ALT, ADEOS II L0</b> <b>JPL - V0 Migration</b> <b>NSIDC - V0 Migration</b> <b>ORNL</b>	<b>18 GB</b>  <b>5 - 10 GB</b> <b>&lt; 1 GB*</b>  <b>1 GB</b>	<b>70 GB*</b> <b>&lt;&lt;1 GB*</b> <b>5 - 10 GB</b>  <b>47 GB*</b> <b>1 GB</b> <b>139 GB**</b> <b>136 GB</b> <b>10 - 15 GB</b> <b>&lt; 1 GB*</b> <b>10 - 15 GB</b> <b>&lt;&lt; 1 GB</b> <b>&lt; 1 GB</b>

**\*Ingested and stored in the ICLHW HWCI**

**\*\*Ingested by the ICLHW HWCI, stored in DSS repository**

# Changes Since Release B IDR



**New technical baseline (February 1996) -- updated size drivers for Release B  
Identified new interfaces and more detailed interface information (based on meetings with external data providers)**

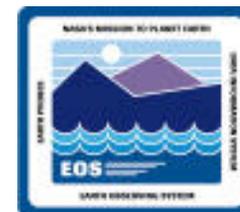
- Revised EDOS interface design
- Established baseline interfaces (e.g., SAGE III)

**Further modularized Ingest software to enhance distributability**

**Integrated with CSS Server Request Framework (SRF) and MSS end-to-end request tracking**

**Completed integration with other infrastructure components (MSS Process Framework, Mode Management)**

# IDR and Workshop Issues Summary



## Status

- **IDR Issue**
  - Performance verification--running performance tests with representative EDOS L0 data volumes, file sizes
- **Ops Workshop Issues**
  - Media check-in/work-off process--agreed to implement media check-in; hooks for bar code reader mechanism included
  - DAN editing--use of standard UNIX editors to edit DAN extracted from data base; HTML Interactive Ingest GUI to allow re-ingest

## Open Issues

- **Retention of data for failed ingest requests**
  - Potential impact on working storage sizing
  - Recommendations provided as an addendum to this CDR

## Next Steps

- **Release B GUI Workshop to review GUI enhancements**
  - Preview as part of this CDR presentation



# Prototypes and Trades

## Status

- **Completed**

- Evaluated AutoSys as a manager of ingest requests; not selected because:
  - a) Ingest request processing is event-driven, not schedule-driven; Ingest is a “short-horizon planner”, not a “scheduler”
  - b) Required use of standalone processes rather than pthreads--performance scalability issues
  - c) Load balancing across processors not implementable due to hardware interface requirements
  - d) Significant additional custom software required to implement

- **Design Decisions**

- Retained Release A custom request management solution; improved distributability

- **In Progress**

- High I/O Performance (EDOS, Landsat-7 interfaces) -- verification of modeling estimates
- Ingest failover -- verification of switchover to backup processor scenario

- **Planned**

- Follow-on to High I/O Performance Prototype
  - \* Data Server benchmarking prior to June, 1996, delta review
  - \* Use of Release A software as testbed during Summer, 1996
  - \* Release B performance testing during Release B I&T phase

# COTS Selections



## COTS Inherited from Release A

- Sybase RDBMS
- RogueWave Class Libraries
- NetScape Commerce HTTP Web Server (shared with DDSRV, with security updates)
- AMASS File Storage Management System
- SGI Challenge L SMPs
- RAID-5 and RAID-1 (for Sybase mirrored transaction log)

## New COTS in Release B

- Illustra DBMS (shared with or reused from Data Server)
- Hardcopy Digitizing/Scanning hardware/software
- EMASS Archive Tape Library (Grau robotics/3590 media)

# COTS Selections (cont.)



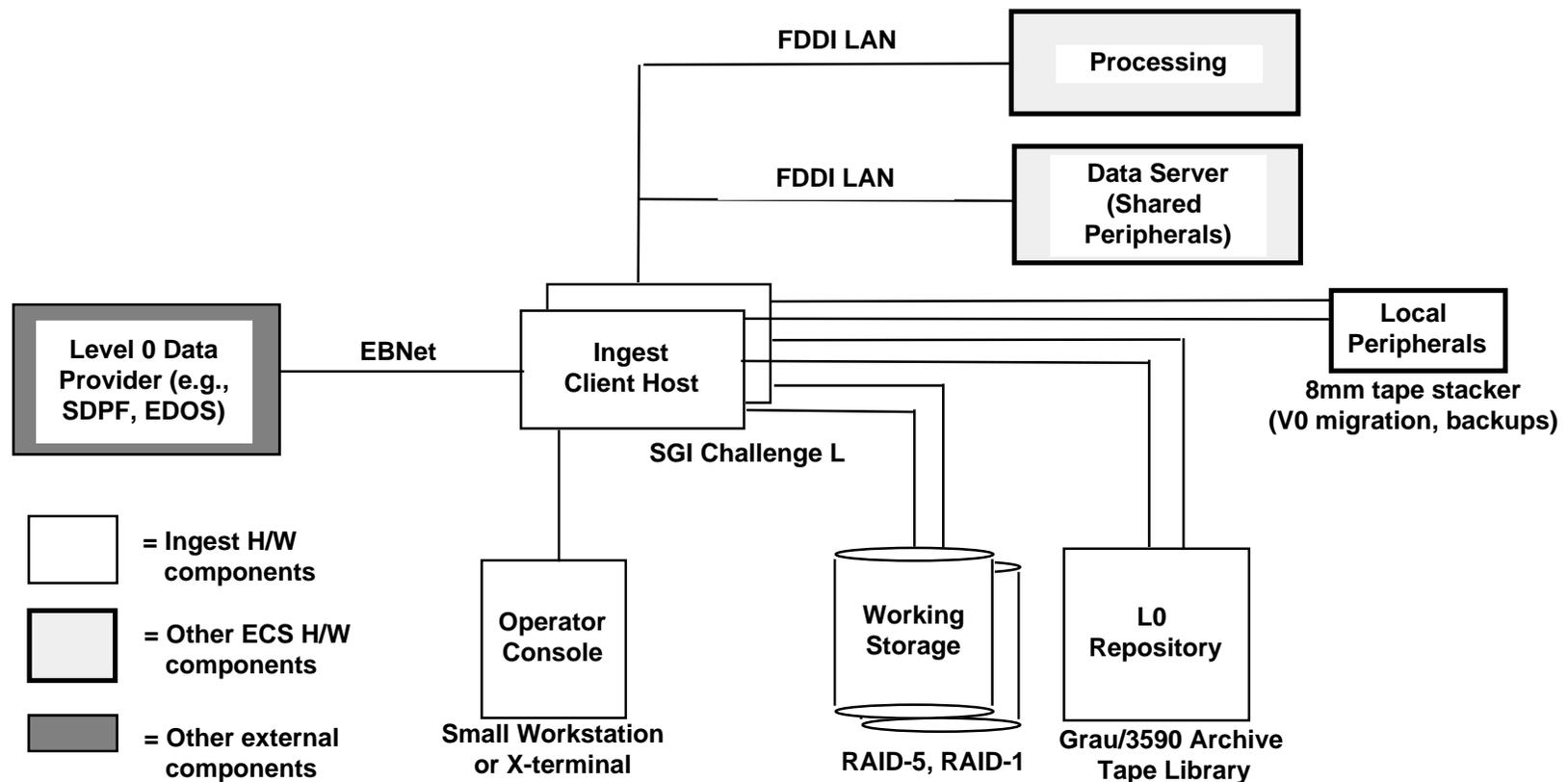
## Status

- Version 0 data migration sizing under way
- Refining L0 data storage requirements

## Next Steps

- Identification of additional V0 hardware requirements
- Final sizing of RAID and Archive Tape Library capacities
- Plan to refine component list prior to June 1996 procurement

# Ingest Client HWCI (ICLHW) Hardware Diagram





# Ingest Client HWCI

## Ingest Client Hosts

- High-reliability, availability
- Sized to support required ingest data I/O, metadata extraction/ checking, preprocessing, and Version 0 data migration support

## Working Storage

- Sized to support high-availability Level 0 data ingest
- 2.5 days worth of storage for support of Level 0 ingest functions (1.25 days for Landsat-7 LOR)

## Level 0 Data Repository

- Provides storage of Level 0 data for c. 1 year (duration dependent on data type)
- Long-term storage of Level 0 data for data without Level 1A counterparts
- Level 0 repository implemented with RAID at small sites, archive tape library at large sites

## Ingest Workstations

- Operations station for support of monitoring and control of ingest

**All critical ingest components are in cross-strapped or redundant configurations to satisfy RMA requirements**



# Ingest Client HWCI Scalability

**Where appropriate, identified Release B-sized hardware for Release A**

- Upgrade of components, as required, may be implemented as described below

## **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/larger capacity arrays

## **Level 0 data repository**

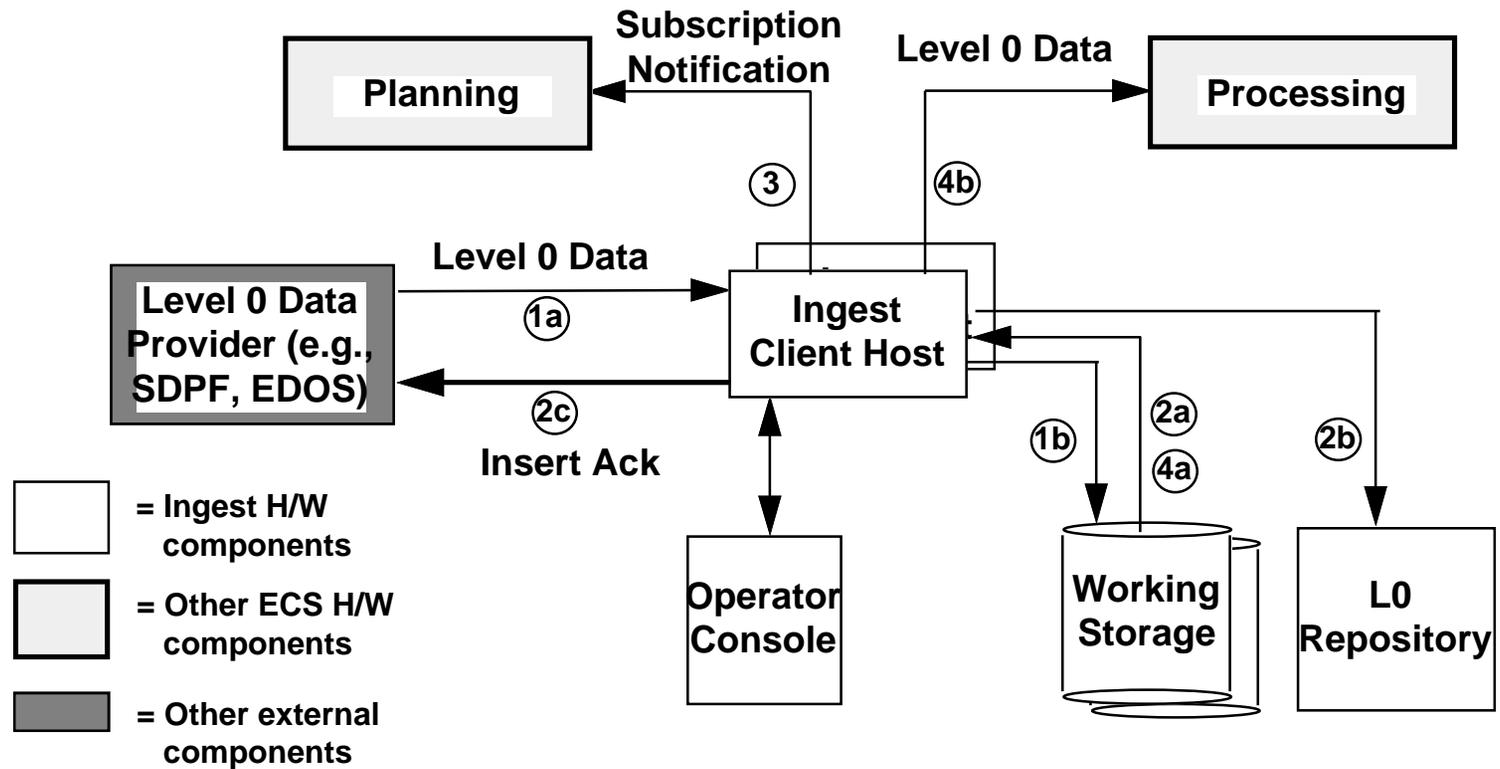
- Addition of tape recorders (for additional I/O throughput)
- Addition of robotics units (for additional data volume)

## **Ingest operator workstations/X-terminals**

- Additional workstations/X-terminals

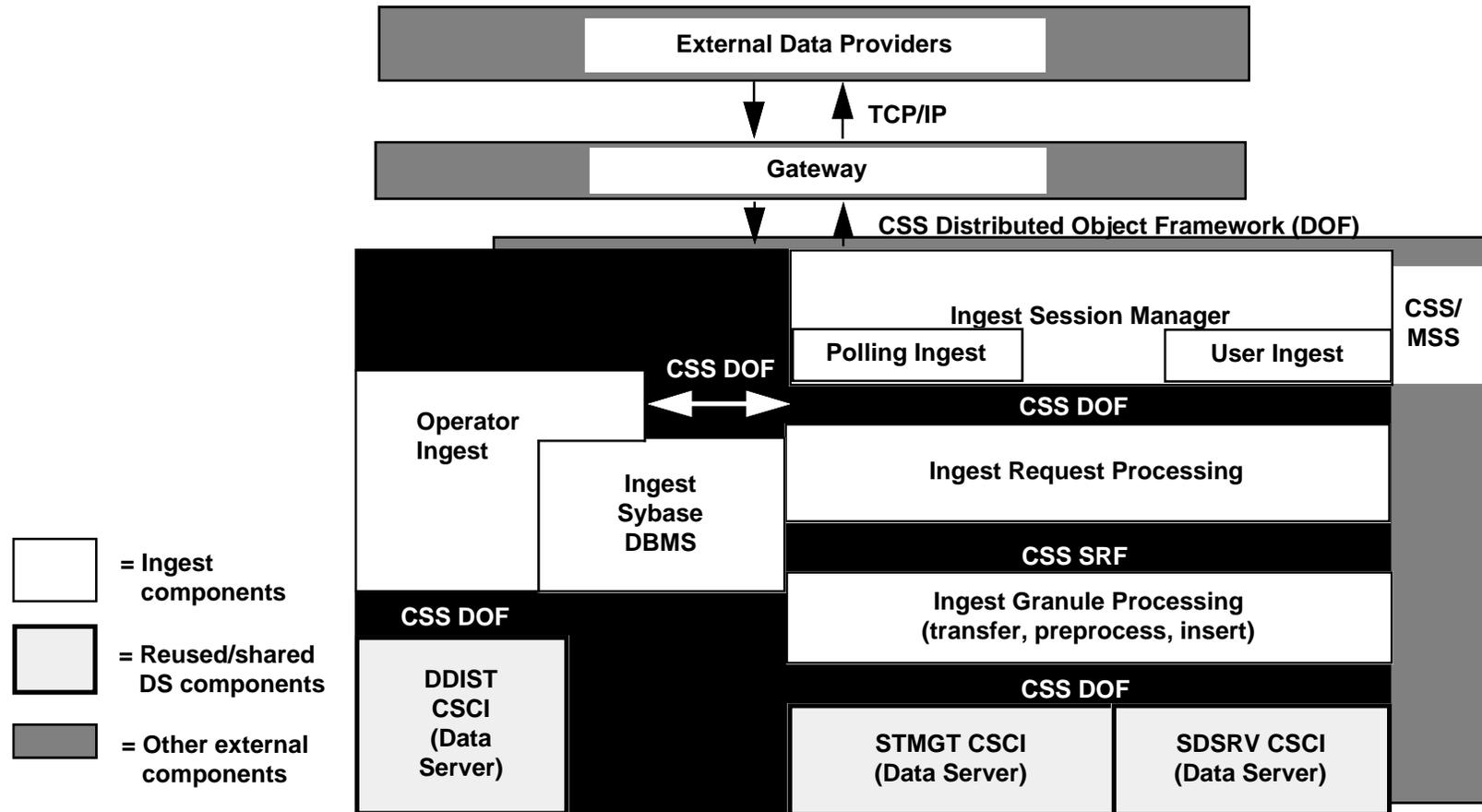


# Ingest Scenario--EDOS L0 Ingest





# Ingest CSCI Architecture

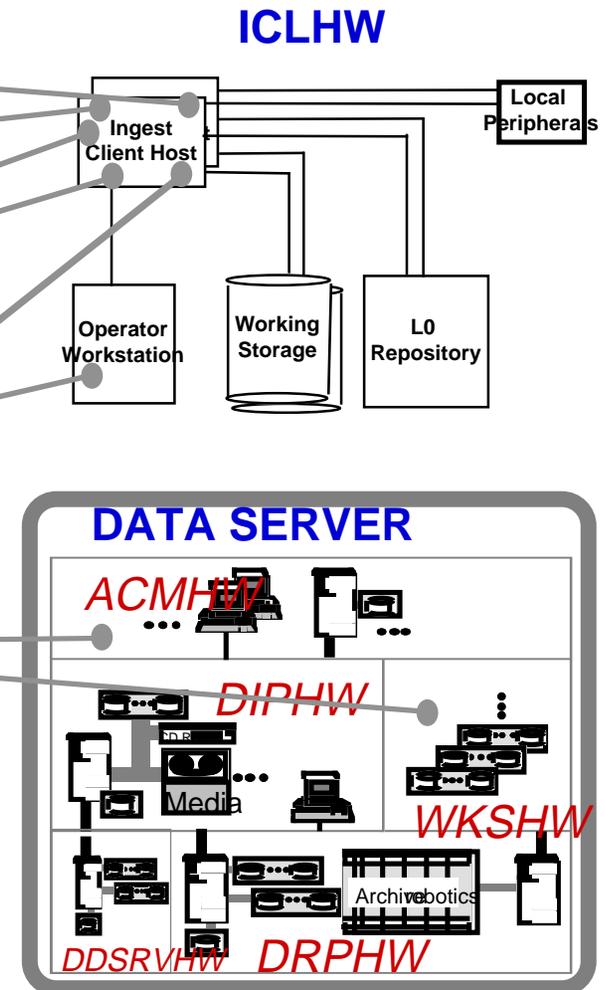




# Ingest Subsystem

## Ingest CSCI Architecture (cont.)

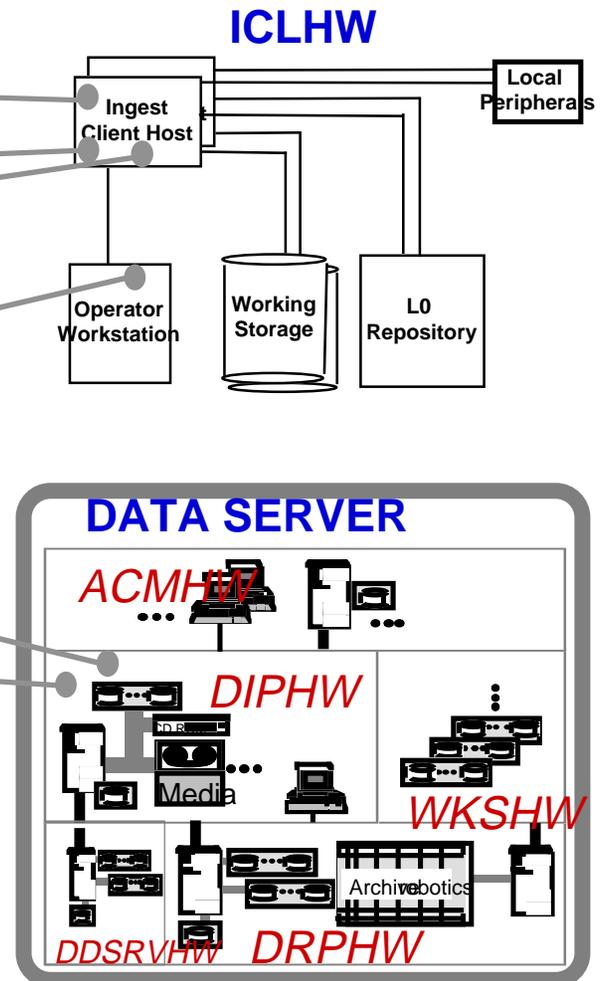
- **Ingest Gateway**
  - Provides hand-shaking for DAN-driven automated ingest (developed by CSS)
- **Ingest Session Manager**
  - Provides hand-shaking for DAN-driven automated ingest
- **Ingest Request Processing Manager**
  - Creates and manages ingest requests
- **Polling Ingest Client**
  - Provides polling interval-driven automated ingest
- **User Network Ingest (GUI)**
  - Provides GUI interactive ingest and status monitoring
- **Ingest Granule Processing Server**
  - Performs granule processing (data transfer, conversions, reformatting, metadata extraction and checking, Data Server interface)



# Ingest Subsystem Ingest CSCI Architecture (cont.)



- **Ingest DBMS (Sybase)**
  - Provides data base storage of Ingest tables
- **STMGT CSCI**
  - Manages Level 0 data repository--reuse from Data Server
- **SDSRV CSCI**
  - Provides data dictionary services and insert service--reuse from Data Server
- **Operator Ingest (GUI)**
  - Provides GUI monitoring and control of ingest processing
  - Media ingest interface
- **DDIST CSCI**
  - Provides media (e.g., 8mm tape) access--share with Data Server





# Other Topics to be Discussed

**GUI interface (preview prior to Release B GUI workshop)**

**Ingest CSC review**

**Key Ingest scenarios**

**Object model/event traces/PDL**

# 305 Errata



**Minor corrections have been made to OMT diagrams and other design materials**

- **Handouts will be provided at the CDR**