

---

# Ingest Subsystem Release A Detailed Design

**Carey Gire**

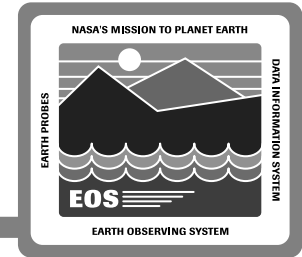
[cgire@eos.hitc.com](mailto:cgire@eos.hitc.com)

---

**ECS Release A SDPS/SCMS Critical Design Review  
15 August 1995**

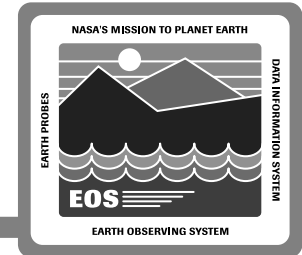
# Ingest Subsystem Outline

---



- **Overview**
- **Design Summary**
- **Changes Since Release A PDR**
- **Standardization of Ingest Interfaces**
- **Ingest Subsystem CSCI (Software)**
- **Ingest Subsystem HWCI (Hardware)**
- **Ingest Requirements for the Community**

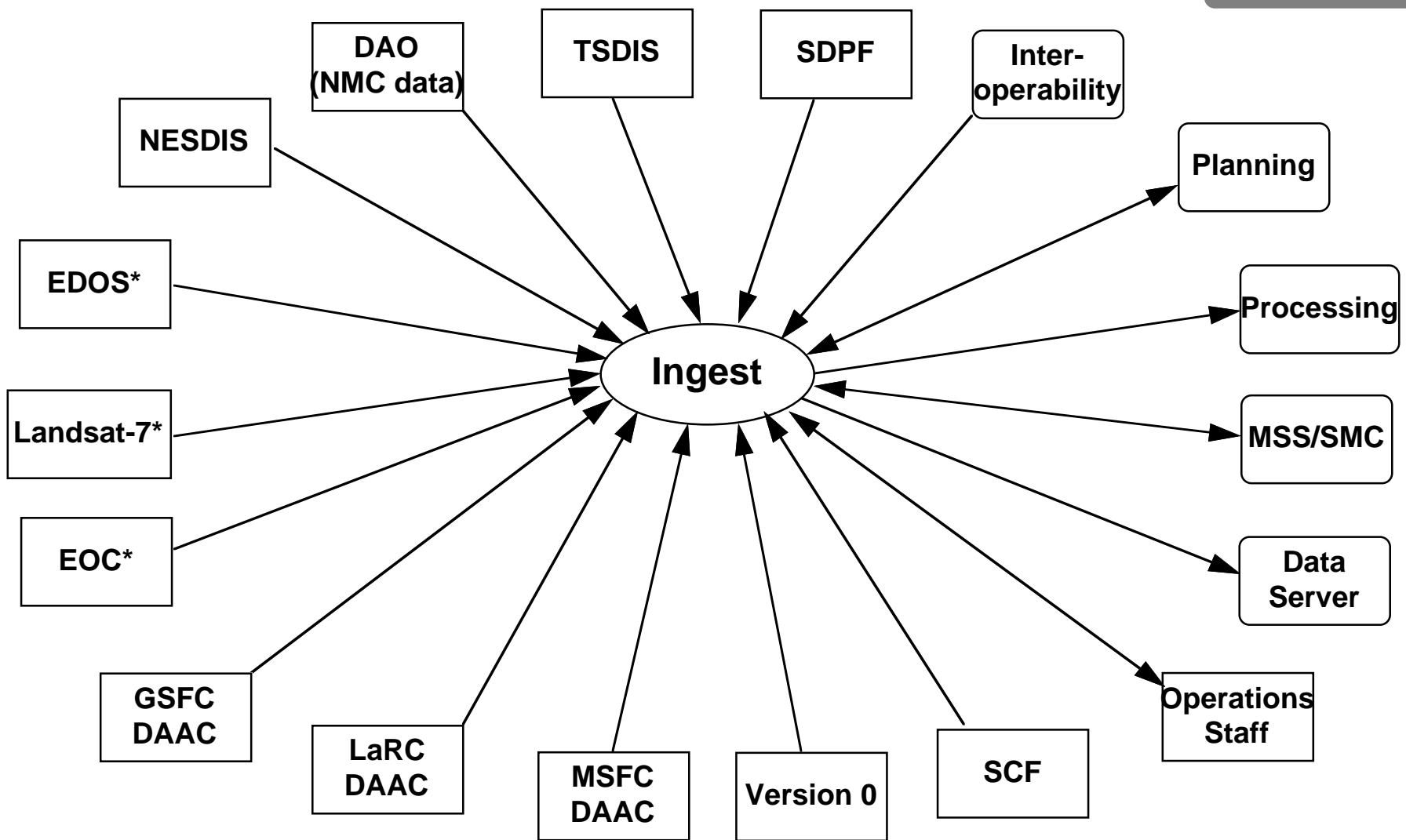
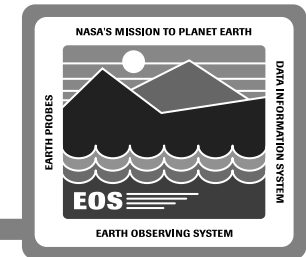
# Ingest Subsystem Overview



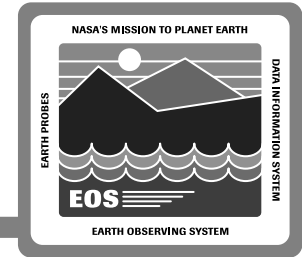
## Release A:

- **Responsible for ingest of data from external data providers (SDPF, TSDIS, DAO, NESDIS, GSFC/LaRC/MSFC DAACs, including V0 data)**
  - **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**
- **Early interface testing for EDOS and Landsat-7**
- **Provides interface for 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

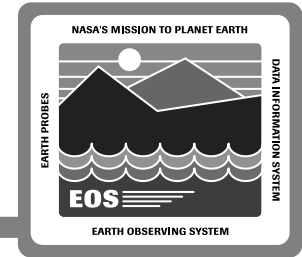


# Ingest Subsystem Design Drivers



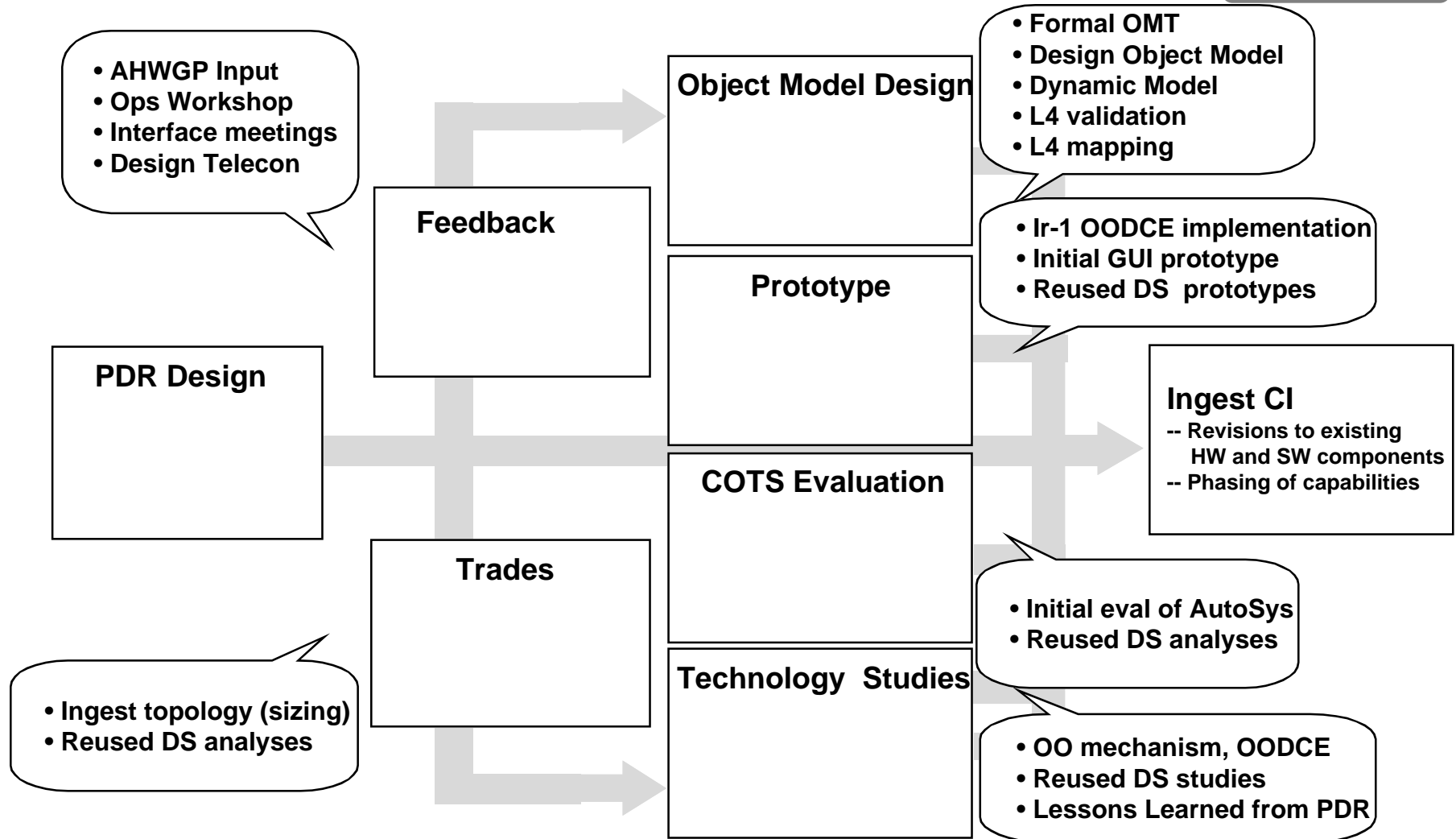
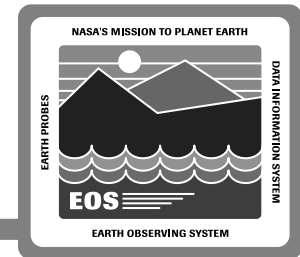
- **Interfaces:**
  - **Very high reliability ingest of Level 0 data (from SDPF)**
  - **1-year Level 0 data storage**
  - **Wide variety of existing interfaces (TSDIS, SDPF, NOAA, etc.)**
  - **Addition of new interfaces and evolvability of existing interfaces**
  - **Version 0 data migration**
- **Operations:**
  - **Maximal automation**
  - **Single point of operations monitoring and control**
  - **Tunable ingest parameters**
- **Exception handling**
  - **Standard CSS/MSS interface to event log with alerts for critical faults**
  - **Component-level recovery after failures**

# Key Ingest Design Concepts

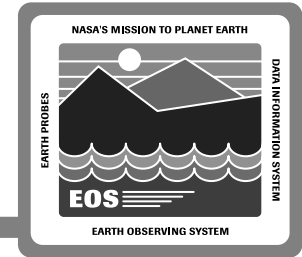


- **Separate, highly-reliable hardware components for Level 0 data ingest**
  - **Ingest “rolling storage” of Level 0 data**
- **Reuse of Data Server software to support Level 0 data storage and retrieval**
- **Standardization of external interface hand-shaking**
  - **Automated, data availability notice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

# Ingest Design Approach



# Changes Since Release A PDR

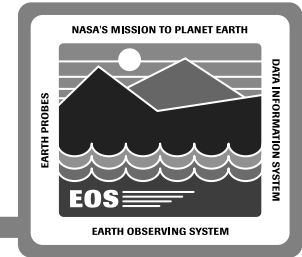


- **Increased daily TRMM housekeeping data volumes**
  - Reviewed use of RAID for Level 0 rolling store
- **Use of OODCE “object factory” paradigm**
  - Standard object creation mechanism
- **Use of DBMS for ingest request browsing and request checkpointing**
  - Store information for duration of request processing; restore request context after system failures
  - Allow browsing of ongoing requests and of request history
- **Final agreement on interface protocols for each external data provider**
  - SDPF, TSDIS, Landsat-7 -- DAN via interprocess communications
  - DAO, NESDIS -- polling for files (with Ingest checklist)
  - EDOS -- polling for Delivery Record file in DAN format



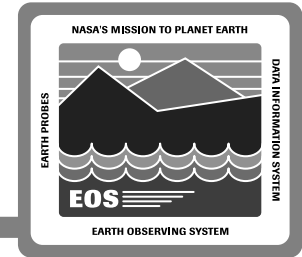
# Changes Since Release A PDR (cont.)

---



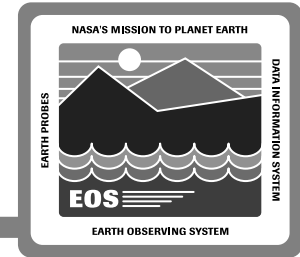
- **HTML user ingest (initially for SCF software delivery package ingest)**
  - **Consistent with proposed forms-based user interface for Release B**
  - **Retain Motif/X-Windows for operator GUI**
- **Implemented custom Ingest Request Processing components**
  - **Release A requirements are simple; cheaper to implement custom software than more complicated COTS (e.g., AutoSys)**
  - **Plan to revisit COTS solution for Release B**

# Standardization of Ingest Interfaces



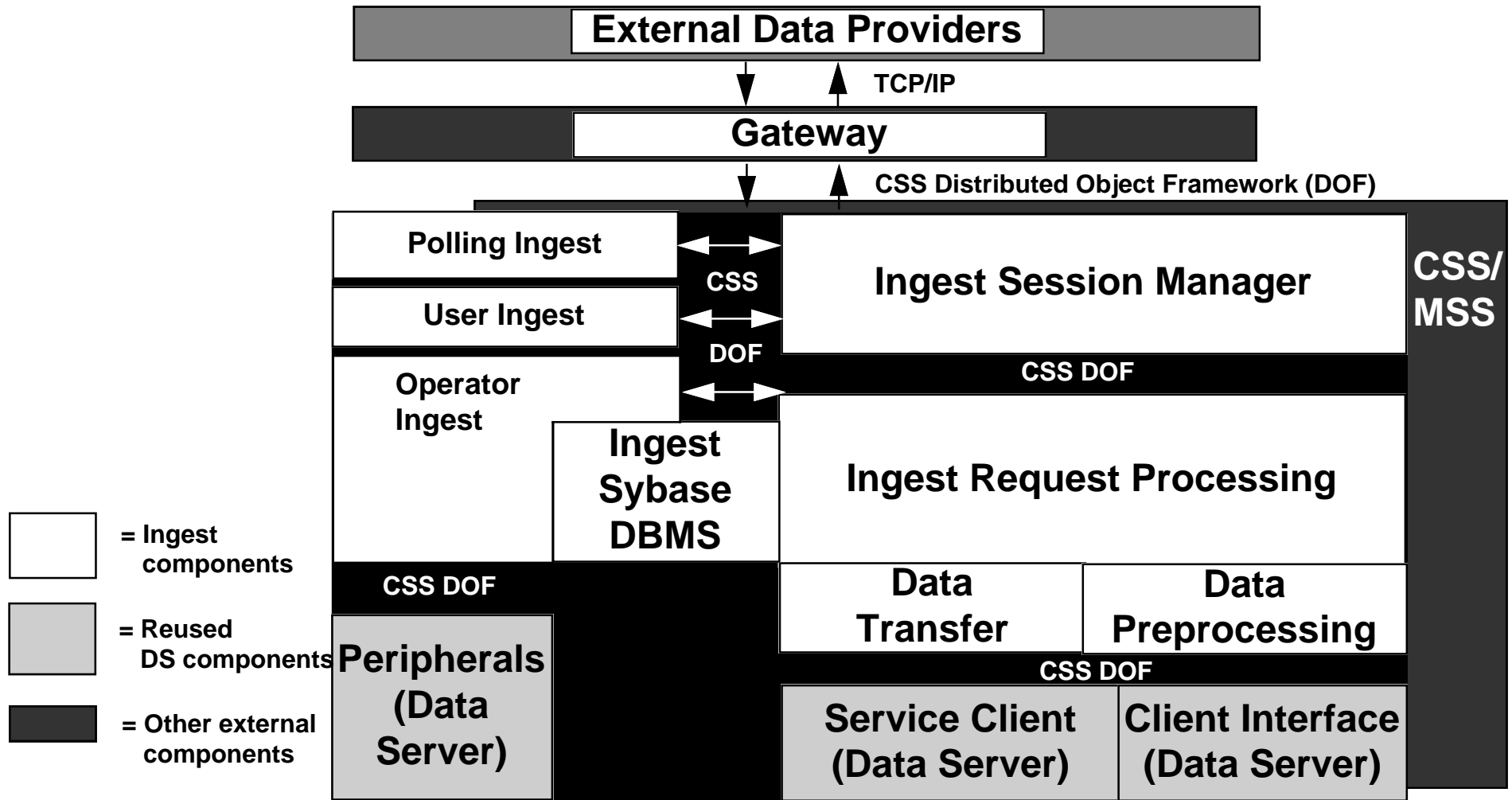
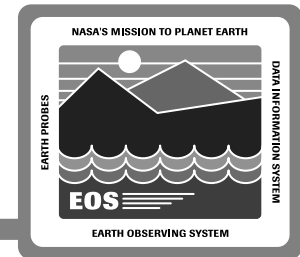
- **Standard applications-level protocol--based on SDPF protocol and CCSDS PVL (Parameter Value Language)**
  - **Data Availability Notice (DAN) provided via interprocess communication or in a file (Delivery Record)**
  - **DAN requires request, data type, and file type descriptors**
- **Standard hard media ingest mechanism**
  - **DAN (stored in a Delivery Record file) used for media ingest as well**
  - **ANSI-standard 8mm tape (tar or cpio format) as default media type (other standard media types added in Release B)**
- **Full Data Server services available if data created using existing HDF-EOS data structures**
  - **ECS services may be limited for native data formats**

# Standardization of Ingest Interfaces (cont.)

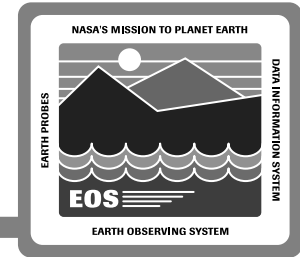


- **Standard metadata storage--a separate metadata file in PVL format is the default standard; alternatively:**
  - **HDF header format**
  - **Byte location/parameter size**
  - **PVL-like format (e.g., specified delimiters and “parameter = value” syntax)**
- **Standard document ingest formats**
  - **HTML with embedded standard keywords**
  - **For other document types the user must supply a separate metadata file in PVL format**

# Ingest CSCI Architecture

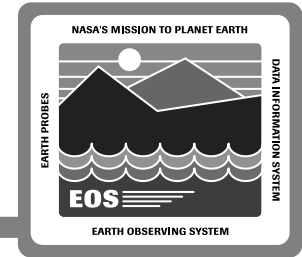


# Ingest Subsystem Ingest CSCI Architecture (cont.)



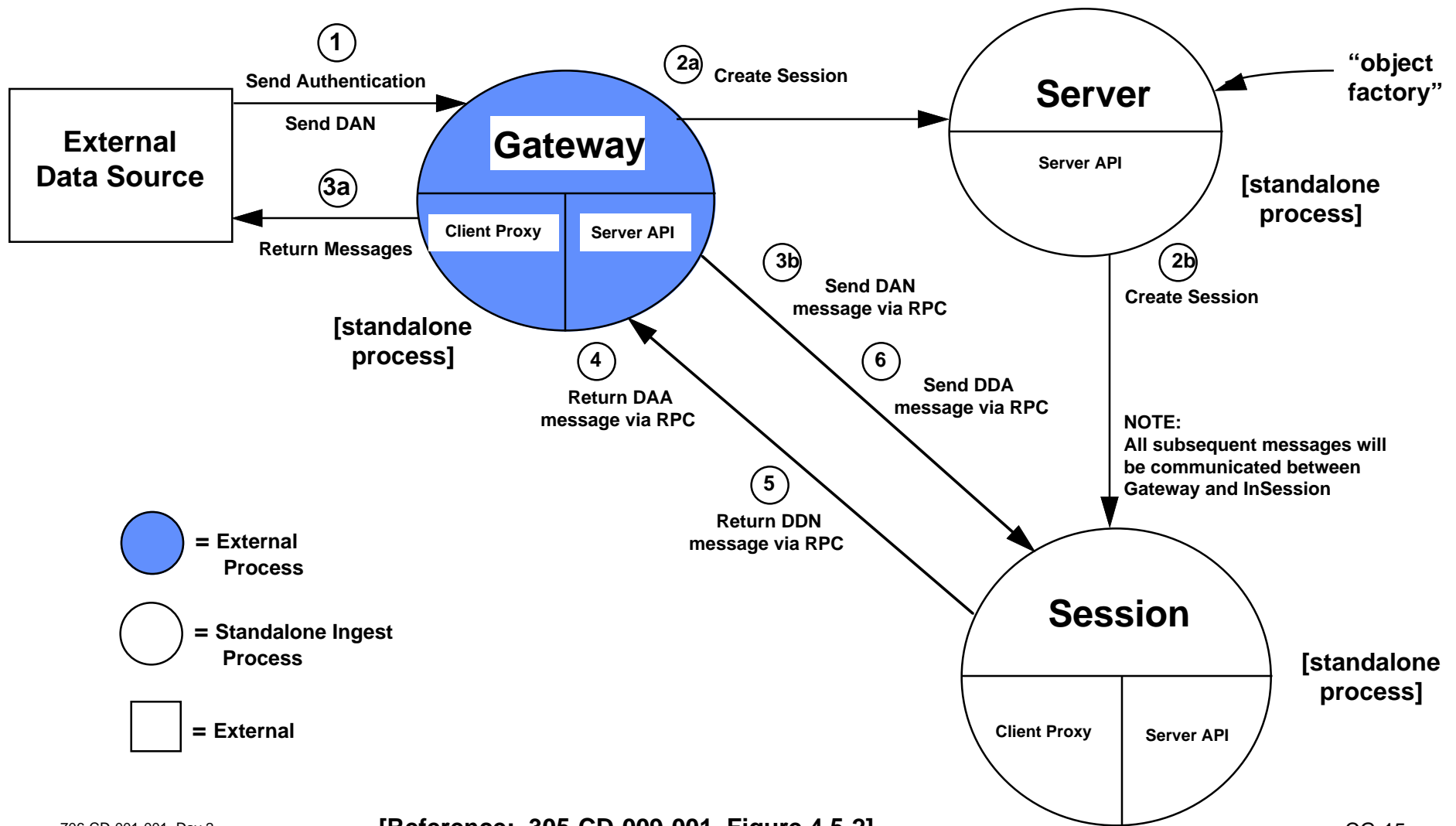
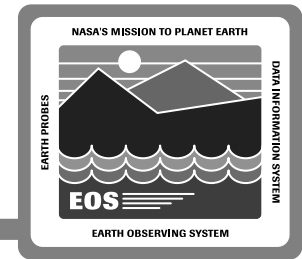
- **Gateway**
  - Provides TCP/IP-to-OODCE translation
- **Ingest Session Manager**
  - Provides hand-shaking for DAN-driven automated ingest
- **Ingest Request Processing Manager**
  - Creates and manages ingest requests
- **Data Transfer Manager**
  - Manages network data transfers
- **Ingest Data Preprocessing**
  - Manages ingest data preprocessing (conversions, reformatting, metadata extraction and checking, Data Server interface)
- **Polling Ingest Client**
  - Provides polling interval-driven automated ingest
- **User Network Ingest (GUI)**
  - Provides GUI interactive ingest and status monitoring

# Ingest Subsystem Ingest CSCI Architecture (cont.)

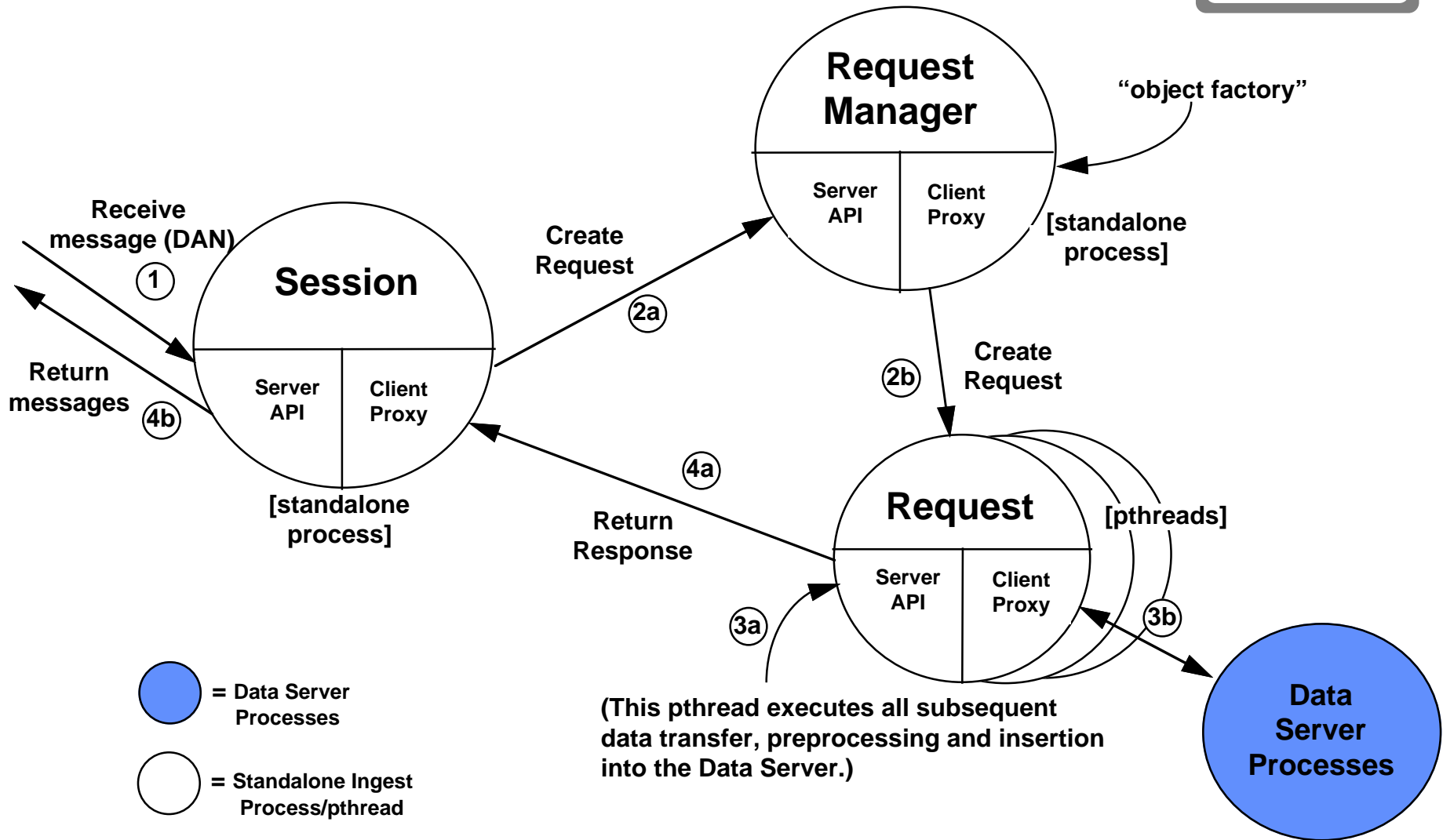
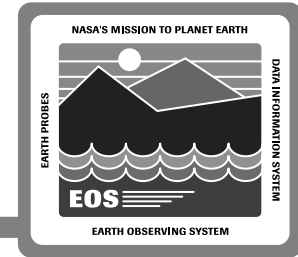


- **Operator Ingest (GUI)**
  - Provides GUI monitoring and control of ingest processing
  - Media ingest interface
- **Ingest DBMS (Sybase)**
  - Provides data base storage of Ingest tables
- **Service Interface**
  - Manages Level 0 data repository--reuse from Data Server
- **Client Interface**
  -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

# Ingest Session Manager Design Scenario

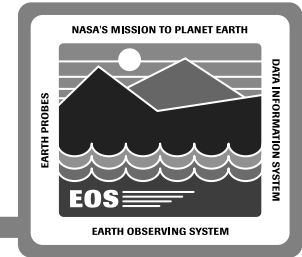


# Ingest Request Processing Components



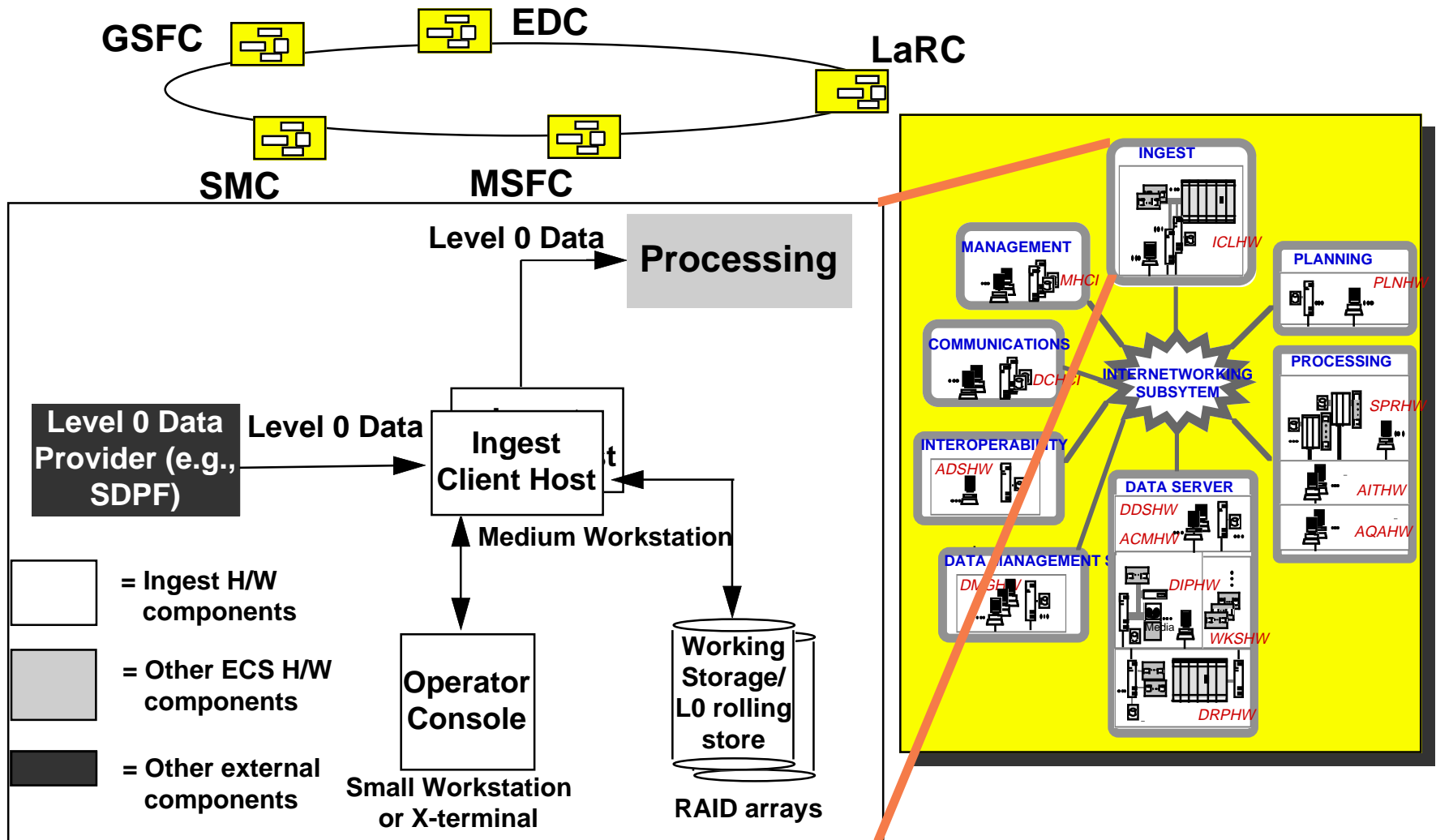
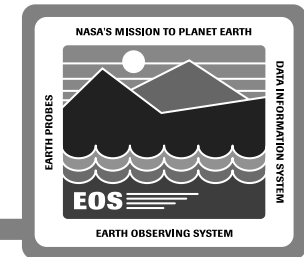


# Operability Features of Ingest

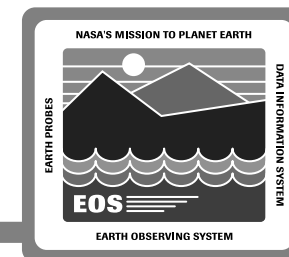


- **Ingest startup/shutdown implemented by MSS life-cycle services**
- **Automatic restart Ingest Session Manager and Ingest Request Processing processes (if failed)**
- **Recover checkpointed session and request states from DBMS**
- **Automatic logging of events related to Ingest; event browsing capability provided by MSS**
- **Standard reports available:**
  - **Request processing summary**
  - **Data set processing summary**

# Ingest Client HWCI (ICLHW) Hardware Diagram



# Ingest Data Size Drivers

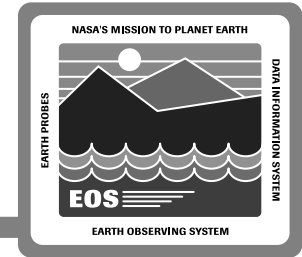


Interface	Volume	Frequency
SDPF - CERES L0*	87 MB	Per day
SDPF - LIS L0*	100 MB	Per day
TRMM - S/C H/K*	162 MB	Per day
TSDIS - All data	53.2 GB	Per day
DAO - NMC data	22 MB	Per day
NESDIS - All data	64 MB	Per week
NOAA (GPCP, GPI)	5.1 MB	Per month
GSFC DAAC (TOMS)	TBR	TBR
LaRC DAAC (SAGE-II, ISSCP)	10.9 MB + TBR (30-40GB)	Per month Per year
MSFC DAAC (SSM/I)	196 MB	Per day
GSFC DAAC (V0)	2000 GB	V0 data migration
LaRC DAAC (V0)	785 GB	V0 data migration
MSFC DAAC (V0)	152 GB	V0 data migration

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

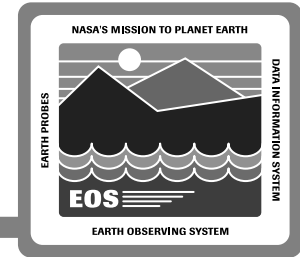
**Note: Only Release A interfaces are shown**

# Ingest Client HWCI



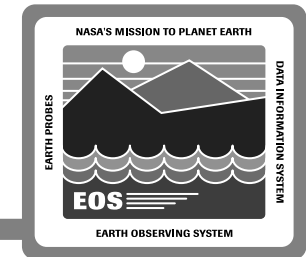
- **Ingest client hosts**
  - **High-reliability, availability**
  - **Sized to support required ingest data I/O, metadata extraction/ checking, preprocessing, and Version 0 data migration**
  - **~50% utilization of CPU and I/O channels**
- **Working storage**
  - **Sized to support high-availability Level 0 data ingest**
  - **>= 2.25 days worth of storage for support of all ingest functions**
- **Level 0 data repository**
  - **Provides Level 0 “rolling storage” (storage of Level 0 data for 1 year)**
  - **Level 0 repository implemented with RAID at Release A**
- **Ingest workstations/X-terminals**
  - **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 DAAC Specific Equipment Purchases



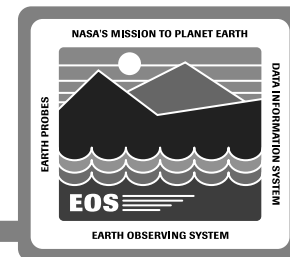
<b>DAAC</b>	<b>Client host</b>	<b>Working Storage (GB)</b>	<b>L0 Rolling Store (GB)</b>
<b>LaRC Calculated</b>	<b>&lt; 0.5</b>	<b>0.625</b>	<b>90.1</b>
<b>LaRC Procured</b>	<b>2</b>	<b>Working and rolling store share 100 GB</b>	
<b>MSFC Calculated</b>	<b>&lt; 0.5</b>	<b>0.660</b>	<b>95.6</b>
<b>MSFC Procured</b>	<b>2</b>	<b>Working and rolling store share 100 GB</b>	
<b>EDC Calculated</b>	<b>0.5</b>	<b>~1 data transmission</b>	<b>0</b>
<b>EDC Procured</b>	<b>1</b>	<b>6</b>	<b>0</b>
<b>GSFC Calculated</b>	<b>0.5</b>	<b>~1 data transmission</b>	<b>0</b>
<b>GSFC Procured</b>	<b>1</b>	<b>6</b>	<b>0</b>

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

# Ingest Subsystem CSCI-to-HWCI Mapping

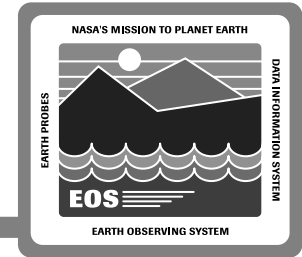


CSCI Component	HWCI Component*
<b>Ingest Session Manager</b> <b>Polling Ingest</b> <b>Operator Ingest</b> <b>User Ingest</b> <b>Ingest Request Processing</b> <b>Ingest Data Transfer</b> <b>Ingest Data Preprocessing</b> <b>Ingest DBMS</b> <b>Data Server STMGT and SDSRV components</b> <b>Peripheral components</b>	<b>ICLHW host</b> <b>ICLHW host</b> <b>ICLHW workstation</b> <b>User Workstation</b> <b>ICLHW host</b> <b>ICLHW host</b> <b>ICLHW host</b> <b>ICLHW host</b> <b>ICLHW host</b>  <b>DIPHW host (Data Server)</b>

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

# Ingest Requirements for the Community

---



- **Final definitions of metadata and valid ranges**
  - **Coordinated with ECS Data Modeling and Data Server teams**
- **Final agreement on early interface activities (EDOS and Landsat-7)**
- **Review of operator/user GUI screens by DAAC and scientist personnel**