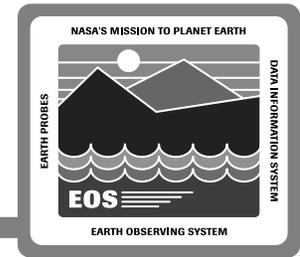


Data Modeling

Denise Heller

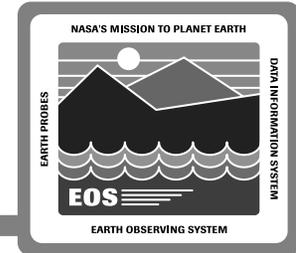
14 February 1995

Agenda



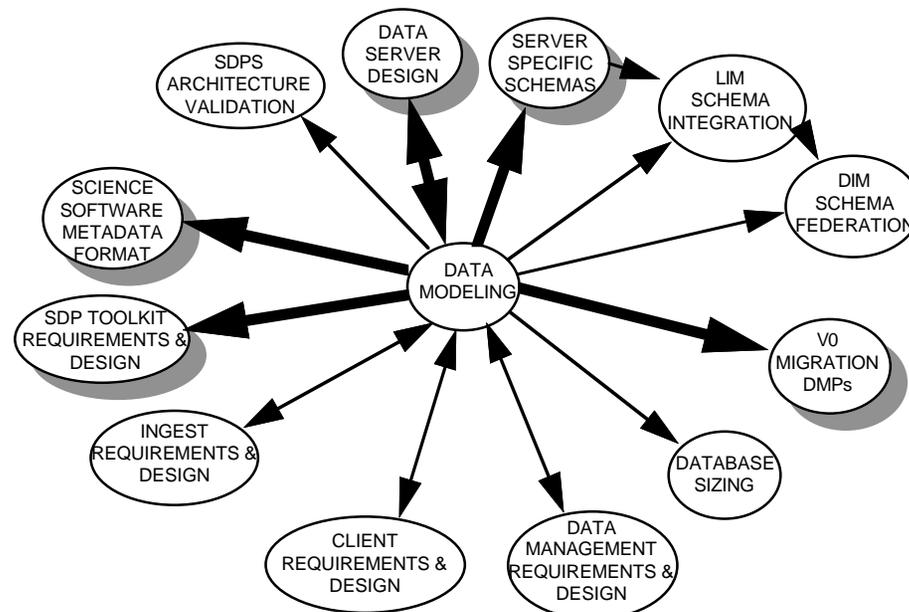
- **Introduction**
- **ECS Core Metadata, proposed standard**
- **Product Specific Metadata**
- **Data Formats**
- **Next Steps**

Data Modeling Context

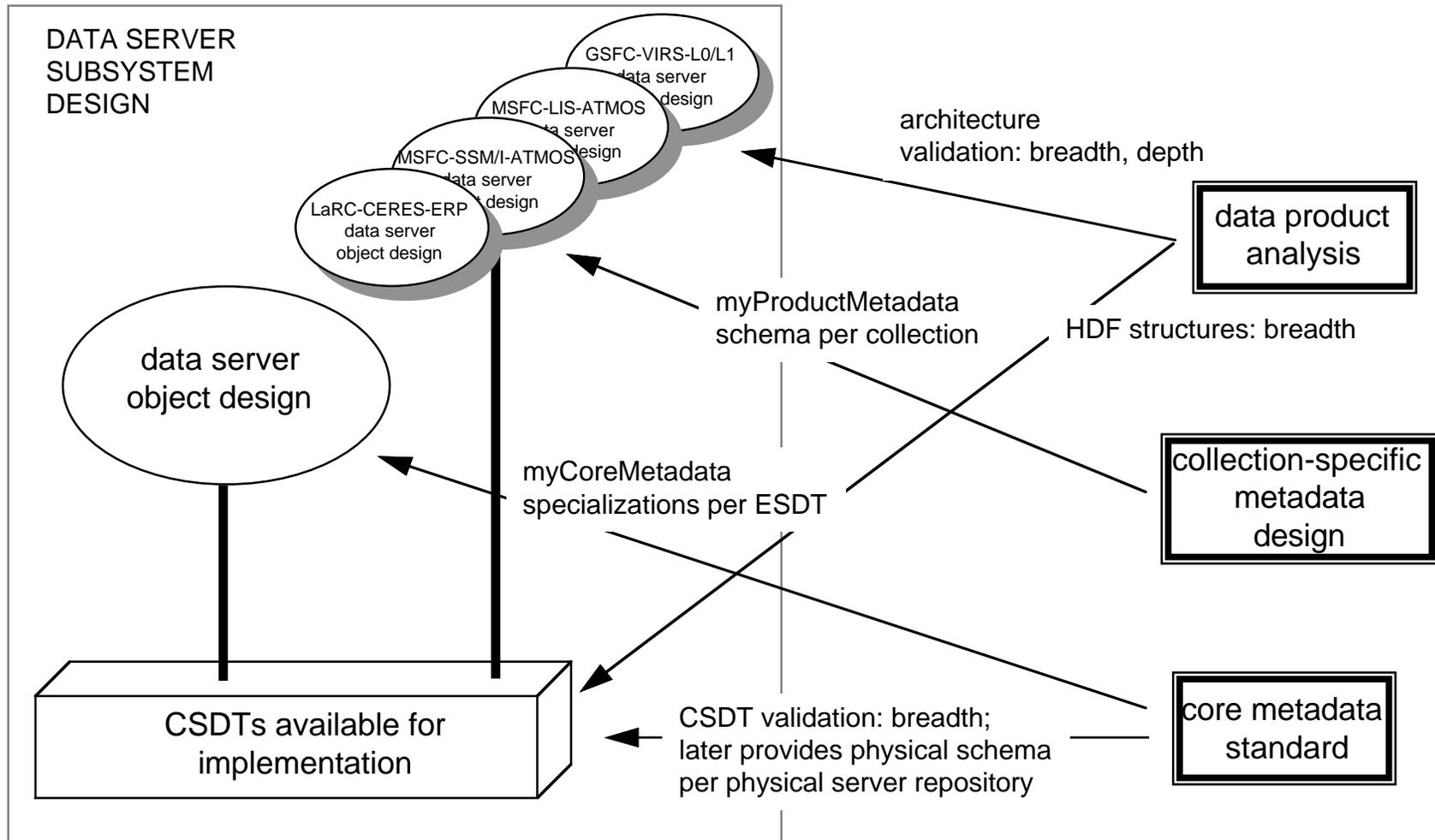
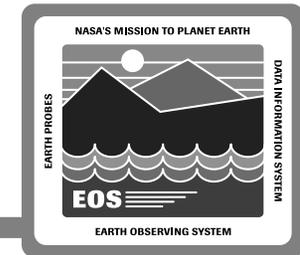


Objective: Provide Standardized Identity of Data

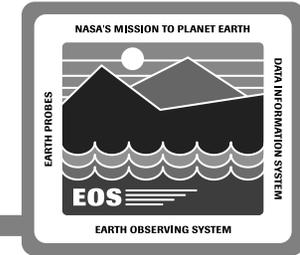
- To system
- To selection process
- To recipients



Data Modeling Context



Core Metadata: Input Sources



User (Producer or Consumer) Forums Attended for input

DOAFT sessions 93
DPFT 4/93
DAAC Managers Meeting, 5/93
Version 0 IMS Workshops, Charlottesville, Hampton 93
SRR 9/93
March Architecture Review
"Mini Science Symposium", 8/93
HDF Data Structure Sessions: ASTER, CERES,
NSCAT/ SeaWinds, MISR, AIRS, (TSDIS, MODIS)
HDF Browse Package Working Group
SDPS Architecture Working Group
PGS Toolkit Reviews, LaRC: 93,94
MODIS MCCST Review Fall 93
MODIS L1A Discussion Spring 94
AVHRR Pathfinder/SeaWiFS lessons learned with HDF

Existing System Analysis

Version 0 Data Dictionary Analysis
(Dataset/DAAC specific data element definition across V0 DAACs)
LaRC IMS Naming Conventions, V0 Naming Standard (JPL/PDS)
EOSDIS Testbed for access to AVHRR/GOES
DIF, Master Directory
TSDIS metadata design
SeaWiFS database design
PLDS schema
GLIS schema
SPOT imagery metadata description
Prototype data: CZCS, URI
OCEANIC interface, Univ of Del Marine Studies

Standards/Reference Models

CEOS Catalogue Subgroup 4/93 Interoperability Recommendations
FGDC Content Standard
PO.DAAC inventory Interface Entry standard SIS-2
TSDIS metadata definition
IEEE Metadata Workshop ref. model draft for Sci and Tech Data
Z39.50 standard for bibliographic reference papers
SAIF: British Columbia, Spatial Archive and Interchange Format
GDAAC Metadata Standard
CIESIN Composition & Development Guidelines for Metadata

Literature Review

Instrument Science Team Meeting Minutes
Instrument ATBDs
JPL User Survey
Modeling in the MTPE Era
Earth System Science
EOS Handbook
Product Catalogs
IWG Meeting Minutes
Version 0 Analysis papers
Sequoia 2000 Database Requirements
RADARSAT issue of Canadian Journal of Remote Sensing

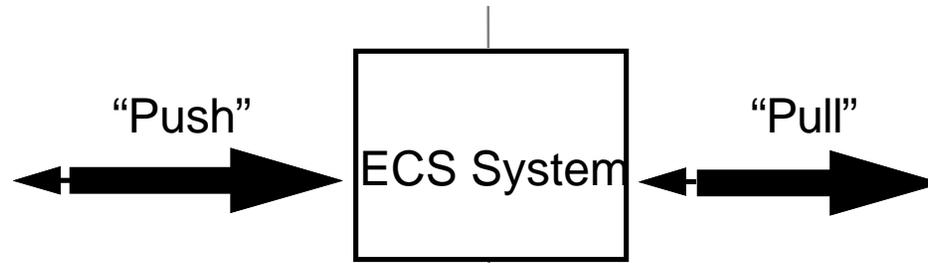
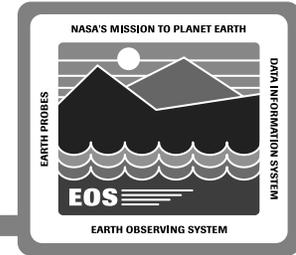
Engineering

SPSO product databases
Data Product Analysis: Inventory, QA/Summ Stats
Data Structure baselines
Browse product baselines
Science Users Guide: Alg Pkg definition

User Community Interviews

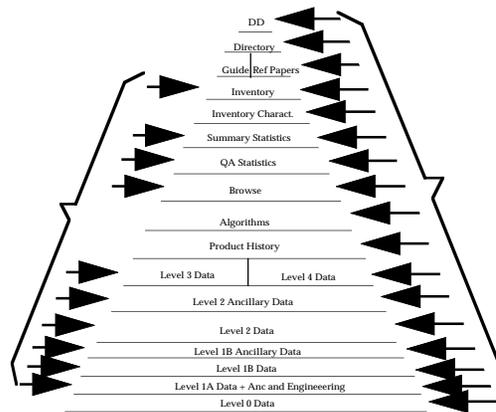
Science Driver Visits Fall '93 by ECS Science Office and
Architecture Team
User Scenario collection visits

Overall Approach: Strawman Core Metadata



What is Planned?

What is desired?



MILESTONES:

12/20/93

3/4/94

4/20

5/31

6/27

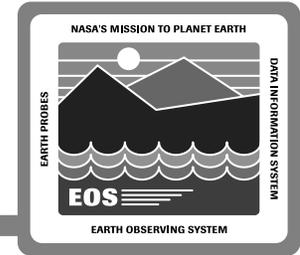
12/94

Taxonomy Phase
(Emphasis on Arch. Support)

Logical Model Phase
(Emphasis on Metadata
Definition)

Align Needs w/
what can be
provided

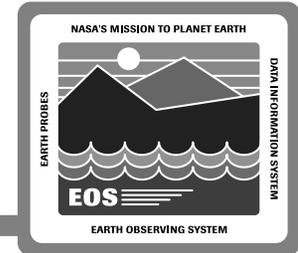
Proposed Core Metadata Standard Agenda



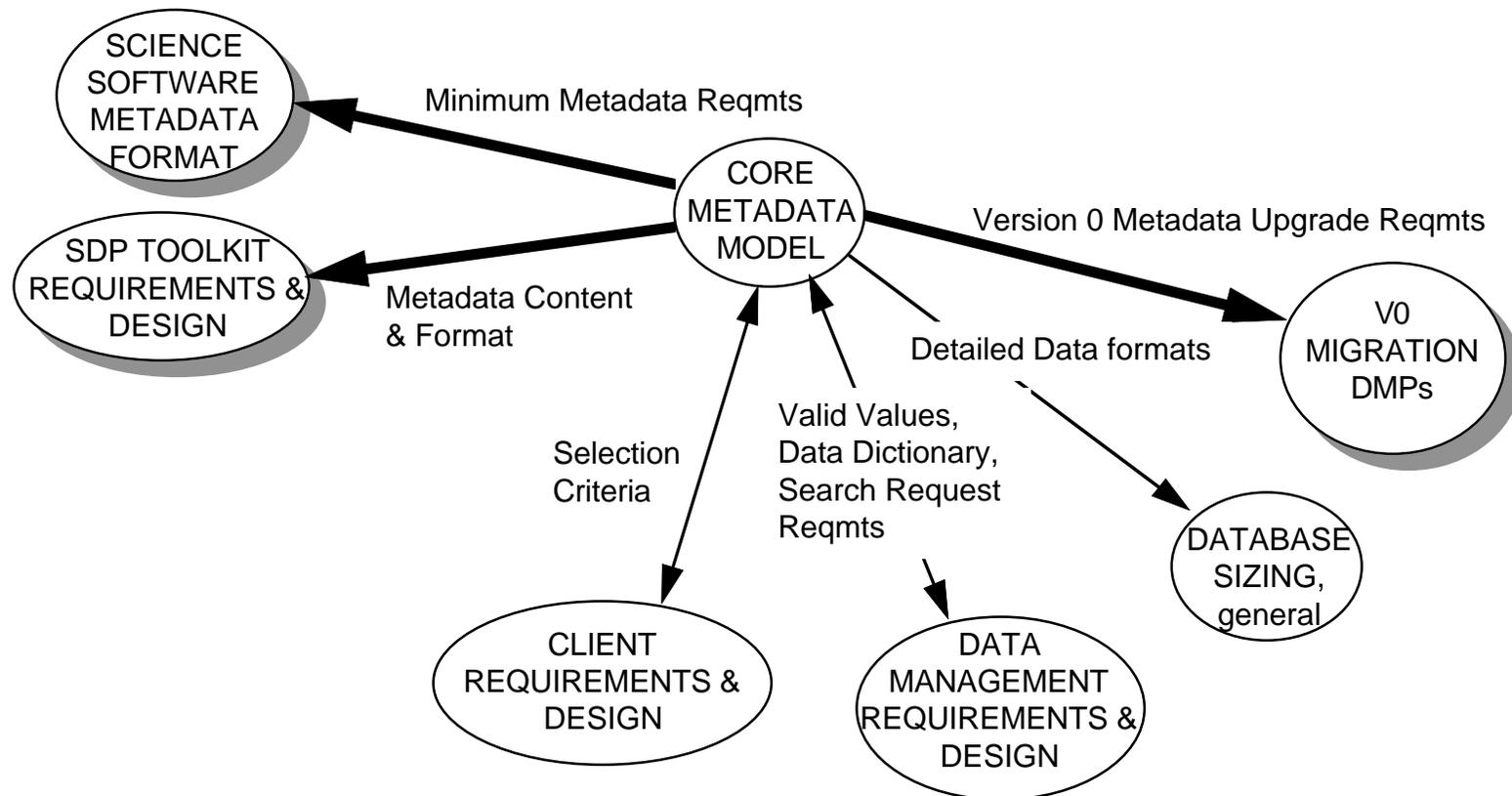
- **Context, Objectives**
- **Core vs. Product-Specific**
- **Basic Components**
- **Ingest vs. Distribution content**

*Document Reference: 420-TP-001-005, <http://edhs1.gsfc.nasa.gov>
-“Proposed ECS Core Metadata Standard”, Release 2.0, 12/94*

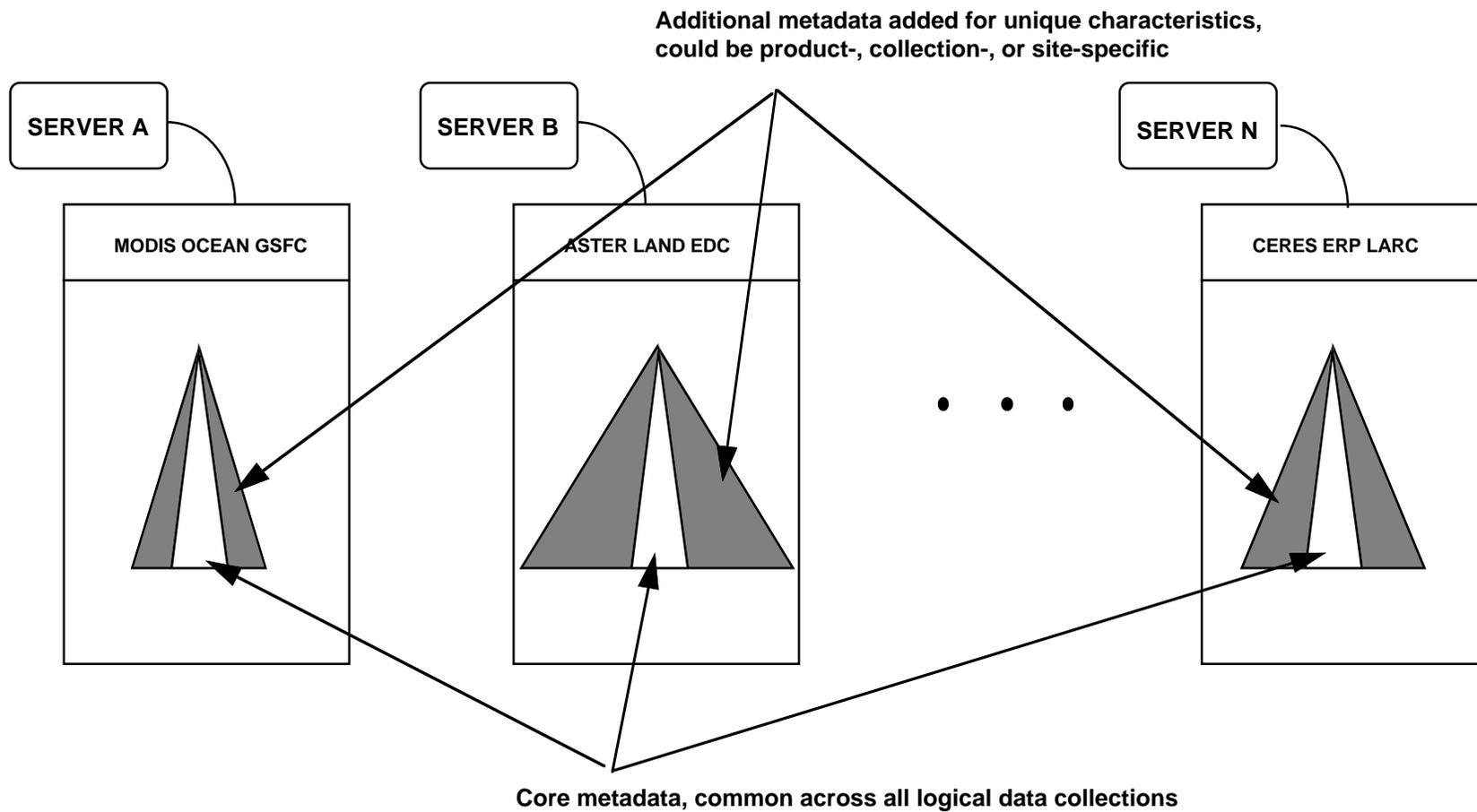
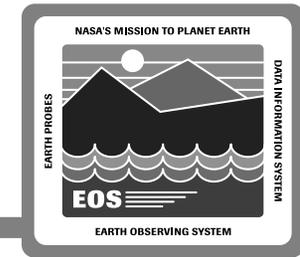
Core Metadata Context



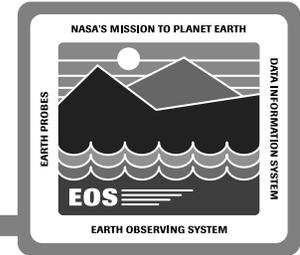
*Plus: FGDC compliance,
GCMD interoperability*



“Core” vs “Product Specific”

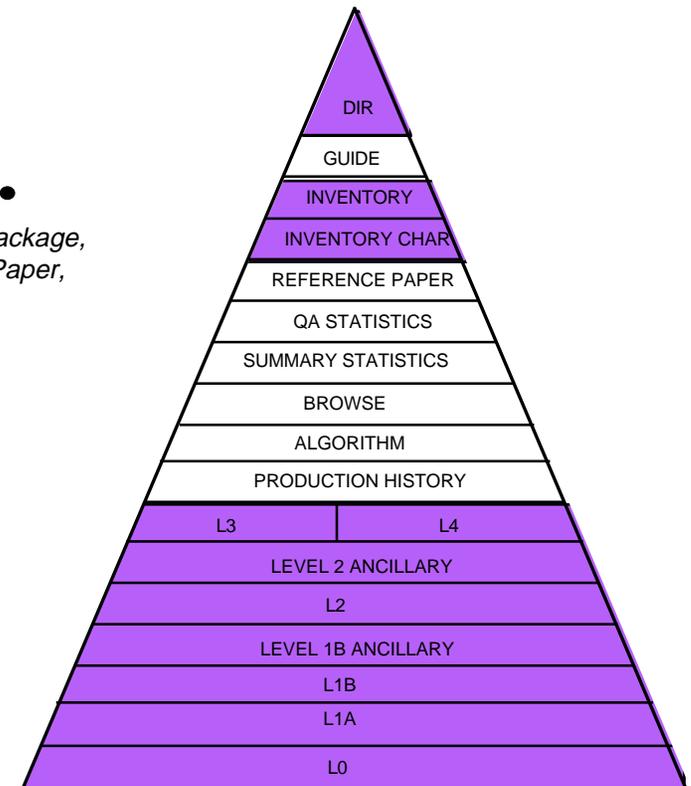
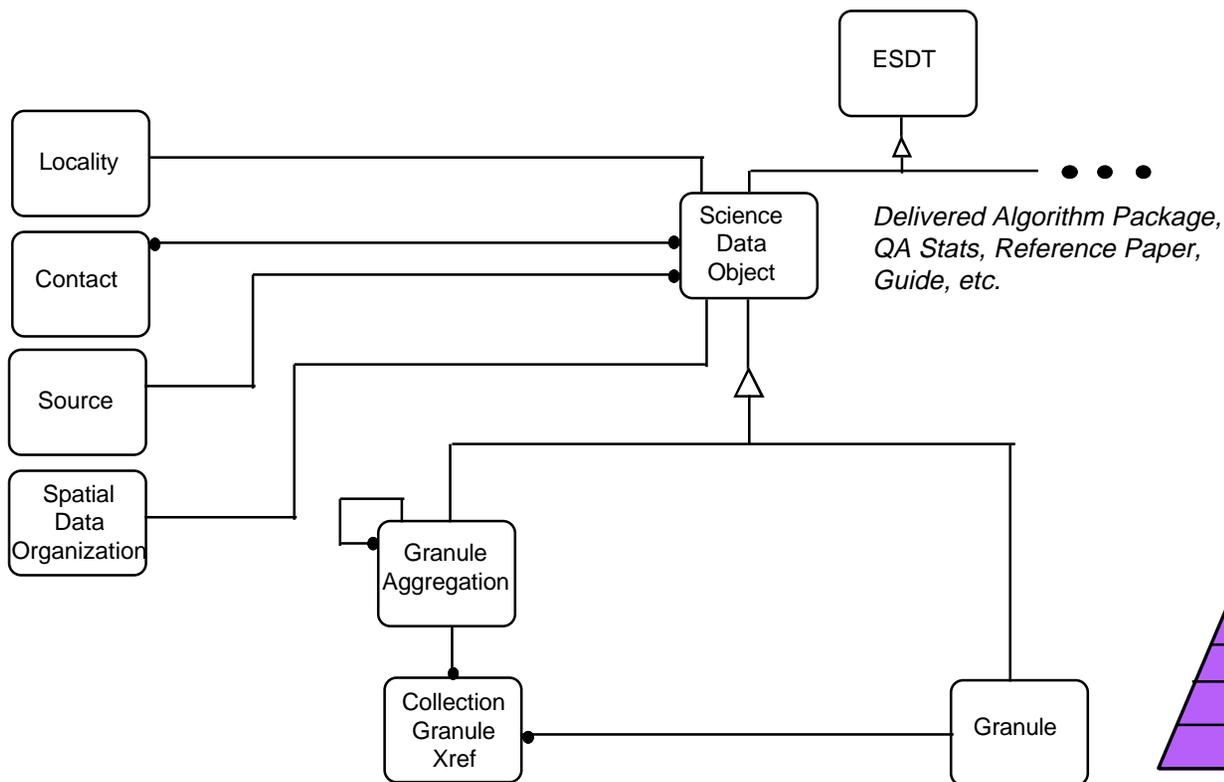
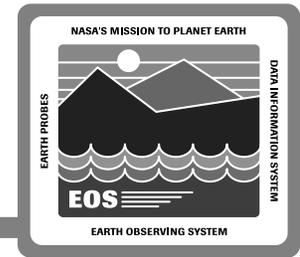


Basic Components

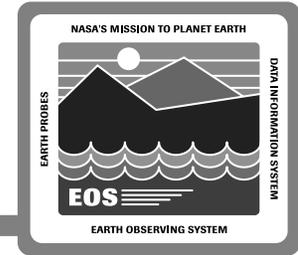


- **Granules and Granule Aggregations**
 - **Granules, Datasets, Products, Supergranules**
 - **Source**
 - **Contact**
 - **Locality: Spatial and Temporal**
- **Documents**
 - **Reference Papers**
 - **Electronic Journals**
 - **Guides**
- **Delivered Algorithm Packages**
- **Other Aids to Selection**
 - **Browse Packages**
 - **QA Stats**
 - **Summary Stats**
 - **Production History**

Objects and Aggregations

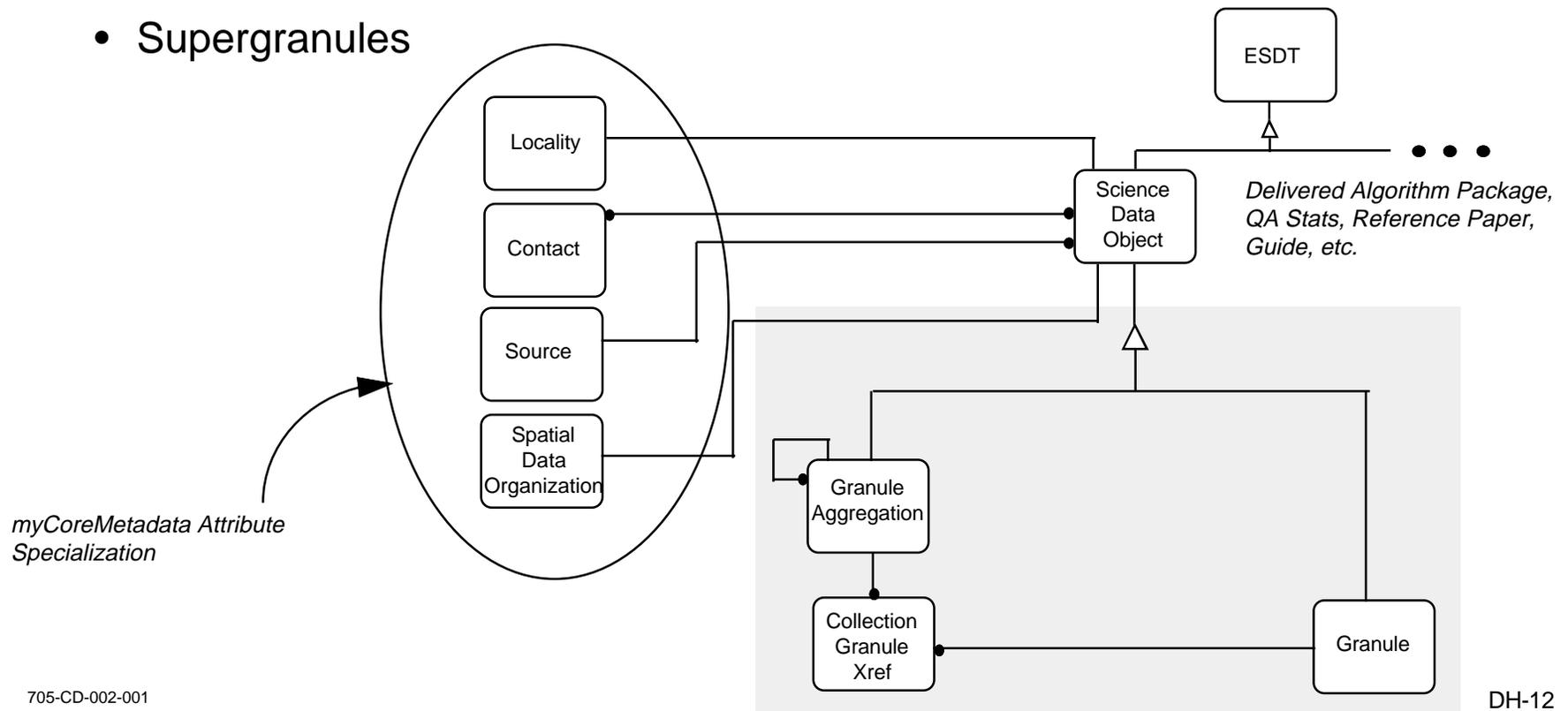


Objects and Aggregations

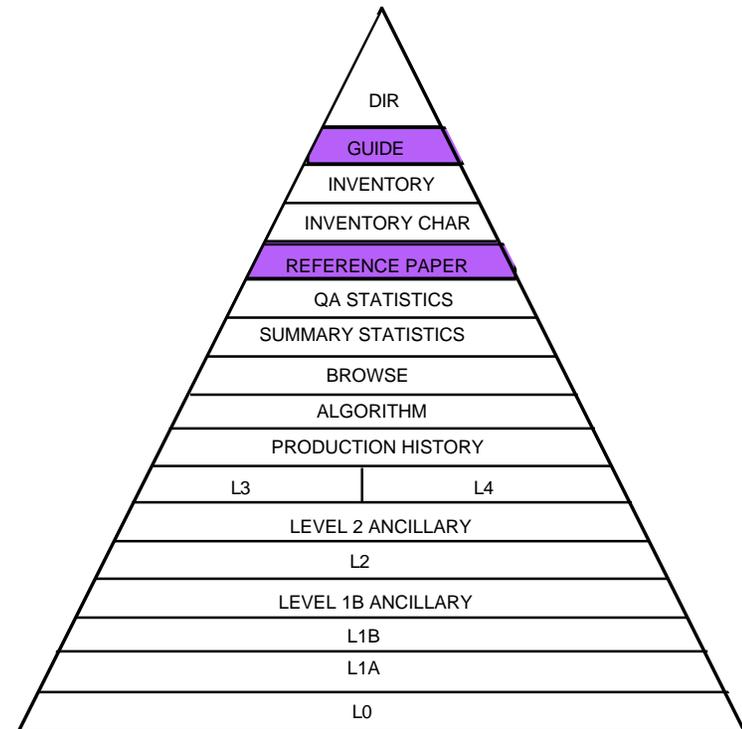
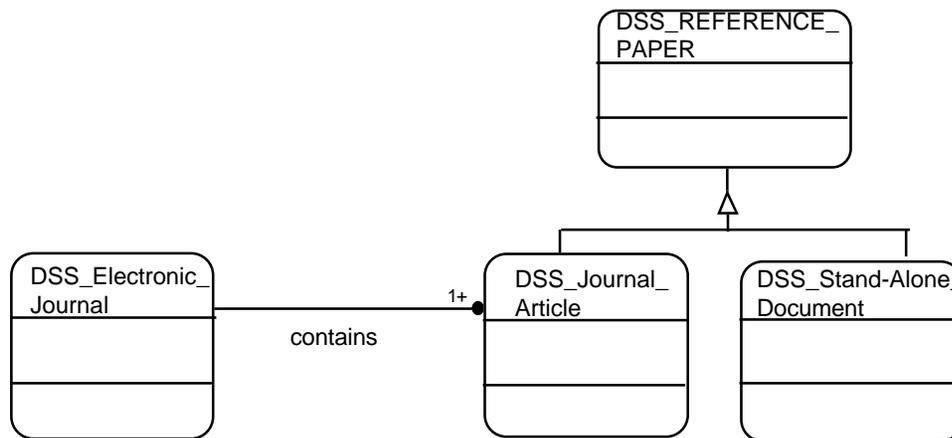
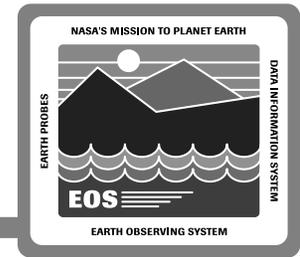


Multiple levels of granule aggregation

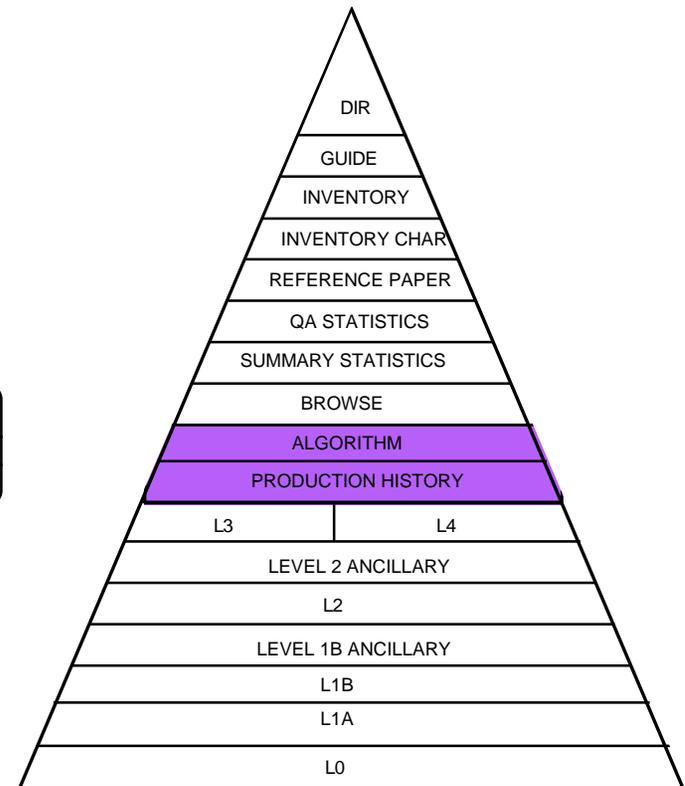
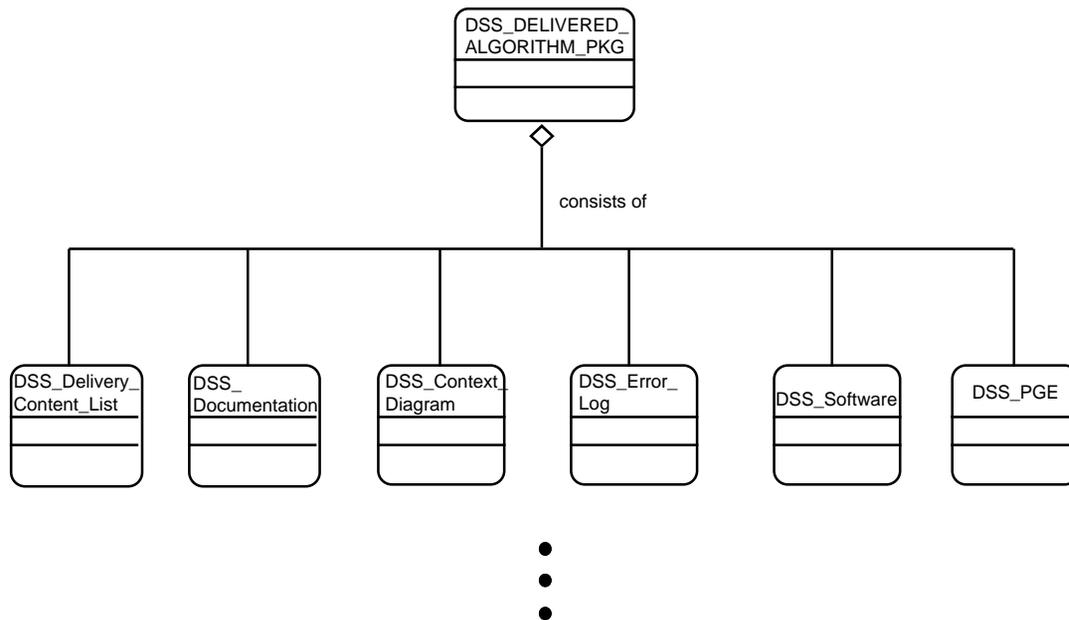
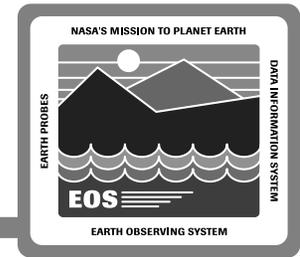
- Products
- Datasets
- Events
- Supergranules



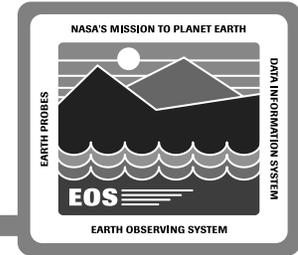
Reference Paper



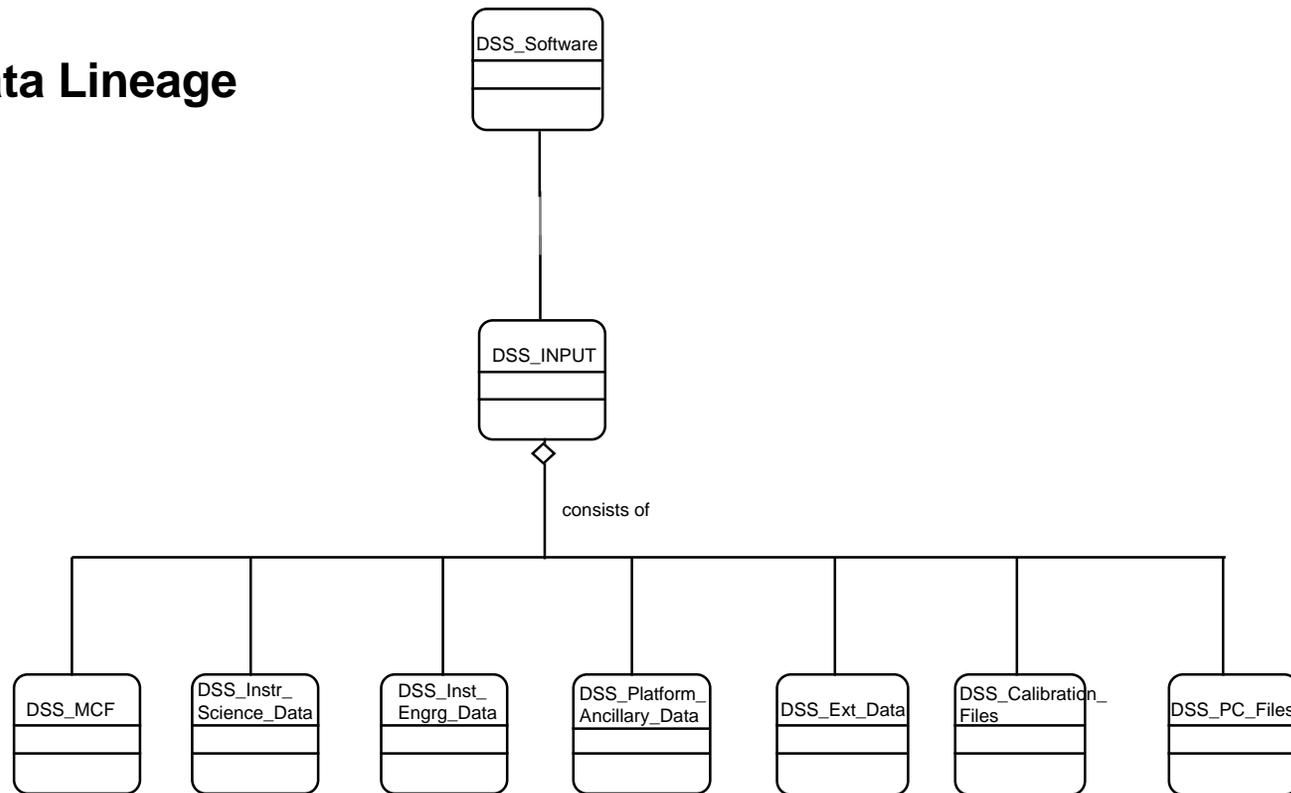
Delivered Algorithm Package



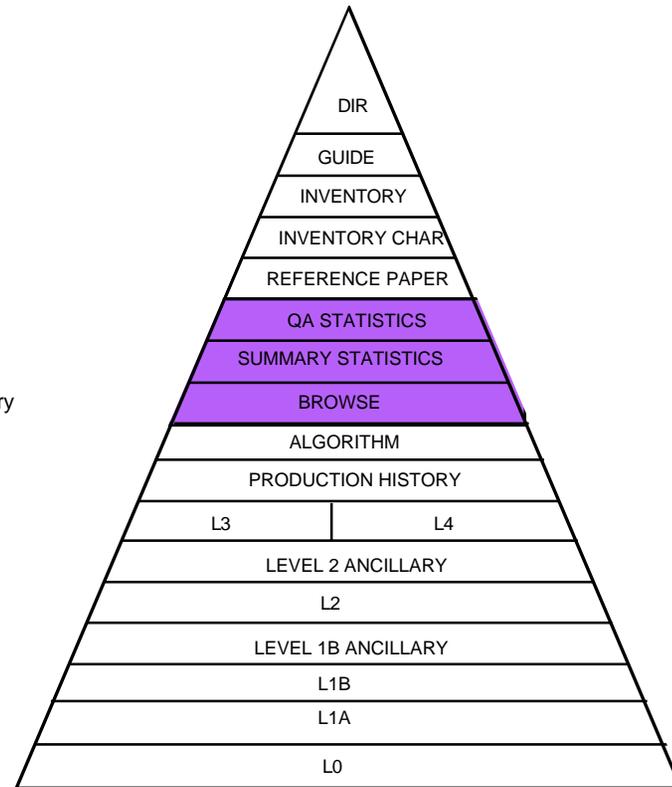
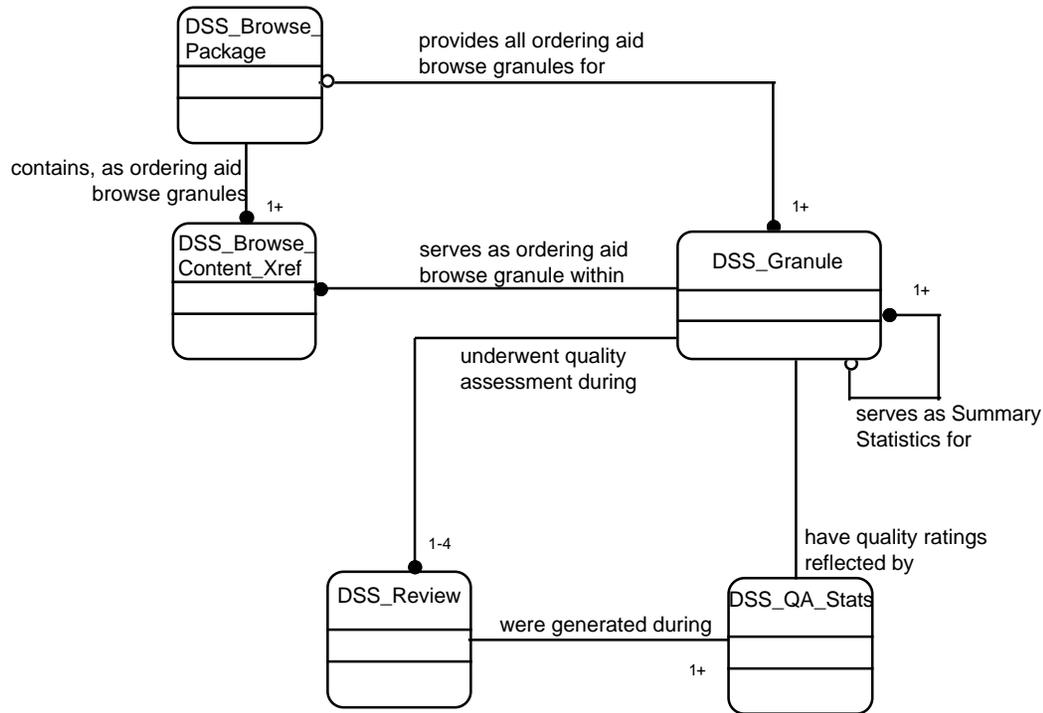
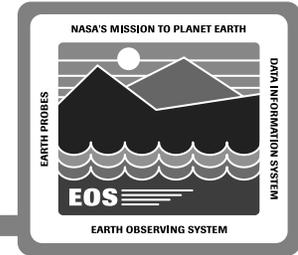
Delivered Algorithm Package



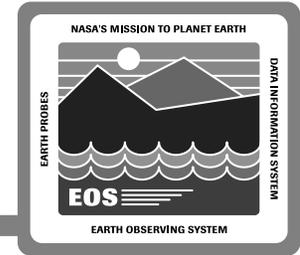
- Captures Data Lineage



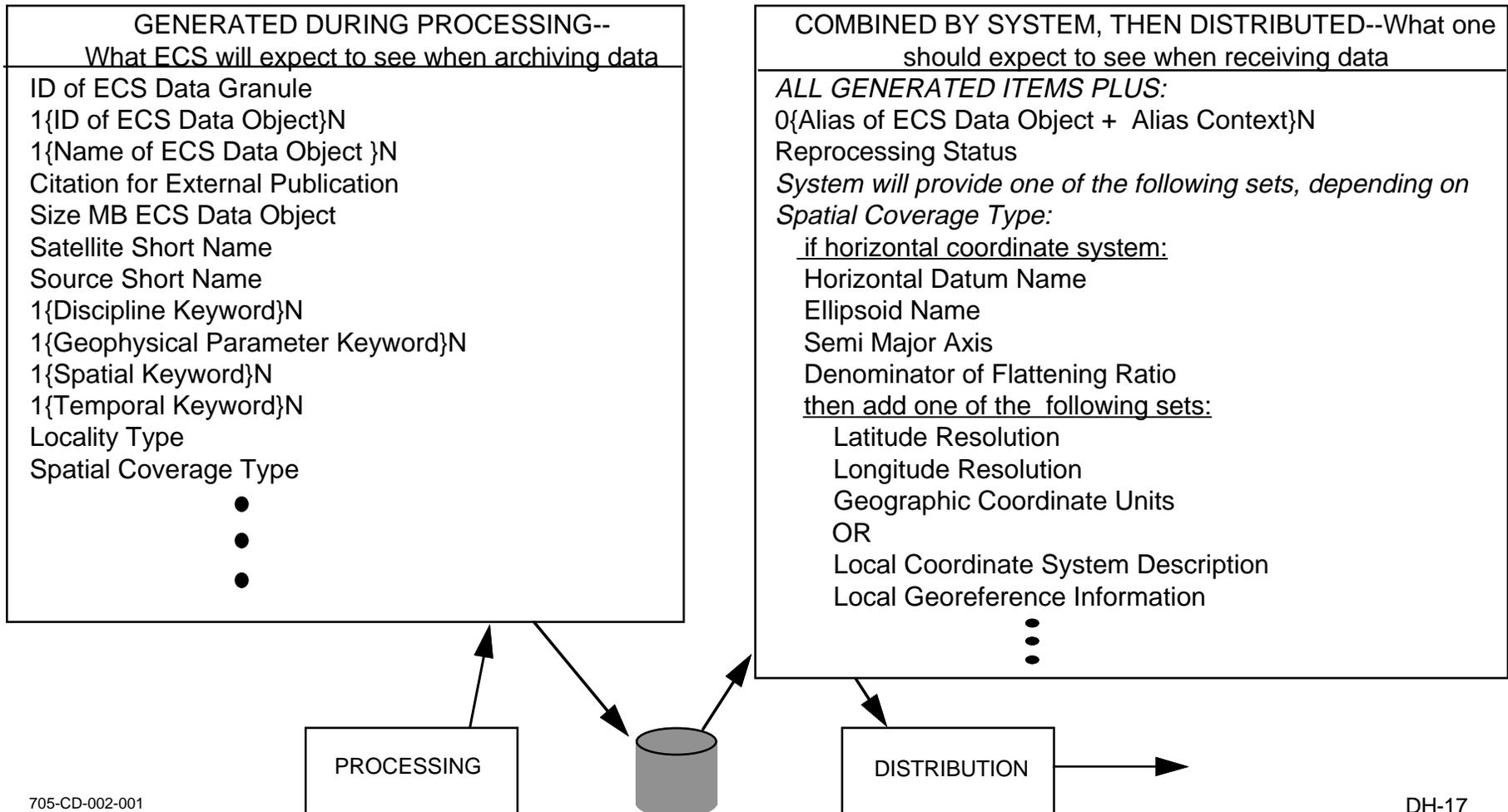
Other Ordering Aids



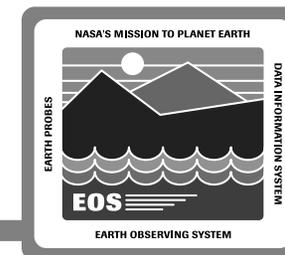
Ingest, Distribution Content



See 420-TP-001, Section 4.8 for complete lists

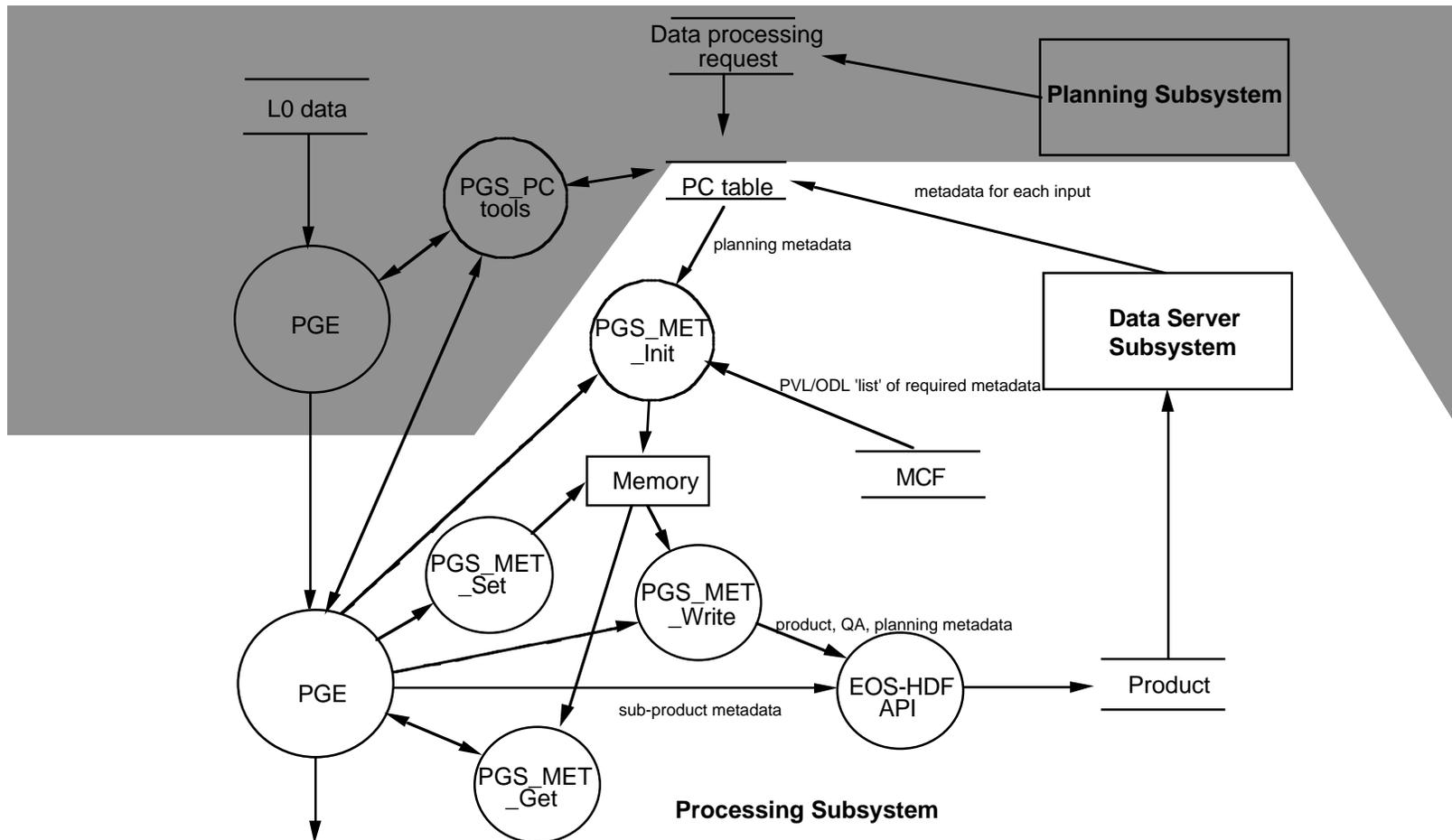
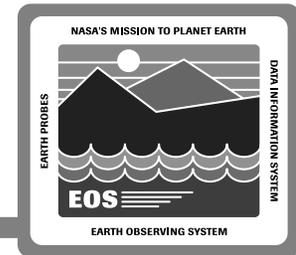


Metadata Generation/Population



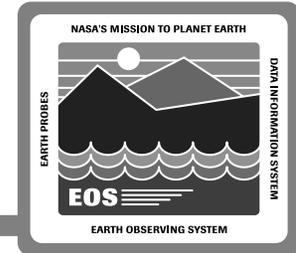
Population Timeframe	Population Source	Data Category
Anytime prior to Initial Product Integration (IPI)	Instrument Teams Stewards	Satellite, Instrument, Project information; Guides; Locality definitions; Processing Level Definitions, List of product-specific attributes, non-standard horizontal or vertical coordinate system definition; non-standard map projection definition. Any information above that cannot be obtained from Instrument Team ; Data Provider Site Guide
As part of Initial Product Integration	Instrument Teams ECS AI&T Personnel	All Delivered Algorithm Package content; Dataset, Algorithm, and Product Subsetting Guides; All Data Collection Object Class attributes; ECS Data Object attributes for this Data Collection; Any aliases of this collection; Spatial and temporal coverage of the collection using Locality object class hierarchy attributes; All Spatial Data Organization attributes for this collection; All points of contact for this collection using Contact object class attributes. Assistance with all items provided by Instrument Teams
Upon update of Algorithms or Alg. Pkg. Content	Instrument Teams	Maturity codes of Del Alg Pkg items; Update to Algorithm Guide to provide rationale for update; Versions, Descriptions, Dates and Effectivities of input elements which may have changed in package
During Production Processing (PP)	Science Software/SDP Toolkit/ECS System Software	All attributes within ECS Data Object, Granule, Browse Package, Browse Content Xref, QA Stats, Review, and Collection Granule Xref object classes; Summary statistics reference; Value attribute within Information Content Attribute object class, if applicable to any Collection to which this granule belongs; Spatial and Temporal Coverage using the attributes of any object class in the Spatial Domain object class hierarchy
Immediately after any QA activity	Personnel or Software performing QA	QA Stats attributes for this data object; Review attributes; Dataset and/or Algorithm Guide updates; Submittal of reference paper
As part of preparation of Version 0 data	Stewards, ECS M&O, Instrument Teams if available	TBS; based on Level of Service; based on Cost Model results; based on FGDC compliance for EOSDIS heritage data
Whenever Users submit papers or QA info	Users with Stewards' Assistance	All Reference paper object class attributes; Association to ECS collection(s)

Metadata Generation Strategy, SDP Toolkit



Document Reference:
193-205-SE1-001, "Science User's Guide and Operations
Procedure Handbook for the ECS Project", 8/93

PVL/ODL Header Format



List of Parameters:

ID_of_ECS_data_granule
1{ID-of_ECS_data_object}N
1{name_of_ECS_data_object}N
citation_for_external_publication
size_MB_ECS_data_object
satellite_short_name
source_short_name
1{discipline_keyword}N
1{geophysical_parameter_keyword}N
1{spatial_keyword}N
1{temporal_keyword}N
locality_type
spatial_coverage_type
spatial_coverage
temporal_range_type
temporal_coverage
future_review_date
science_review_date
science_review_status
1{quality_rating}N
reprocessing_status
PGE_identifier
processing_level_ID
browse_package_reference
1{point_of_contact}N
0{product_specific_data}N



spatial_coverage

Combination of spatial domain attributes, according to processing level and spatial coverage type.

Begins with BEGIN_GROUP = SPATIAL_COVERAGE, ends with END_GROUP followed by ;

- **Level 0/1A, HORIZONTAL**

orbit_model_reference

The reference to the orbital model to be used to calculate the geolocation of this data, in order to determine global spatial extent. Maximum 80 characters; must be enclosed in double quotes followed by ;

[May be changed to ODL file pointer syntax]

3{(bounding_coordinate_name, bounding_coordinate_value)}4

Three or four pairs of values which specify the general spatial extent of the granule coverage, either in terms of North and South Latitude plus East and West Longitude, or in terms of Center Latitude, Center Longitude, and Radius. Global coverage is denoted by four zero values.

Must begin with left curly bracket; each pair must be enclosed within parentheses; commas should separate each parenthesized pair; list must end with right curly bracket followed by ;

bounding_coordinate_name

Name of the coordinate being used to specify general spatial extent or coverage.

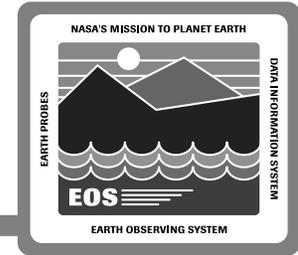
Must be within domain of valid values; must be enclosed in double quotes

bounding_coordinate_value

Value of the associated bounding_coordinate for this locality/coverage.



Product-Specific Metadata Agenda



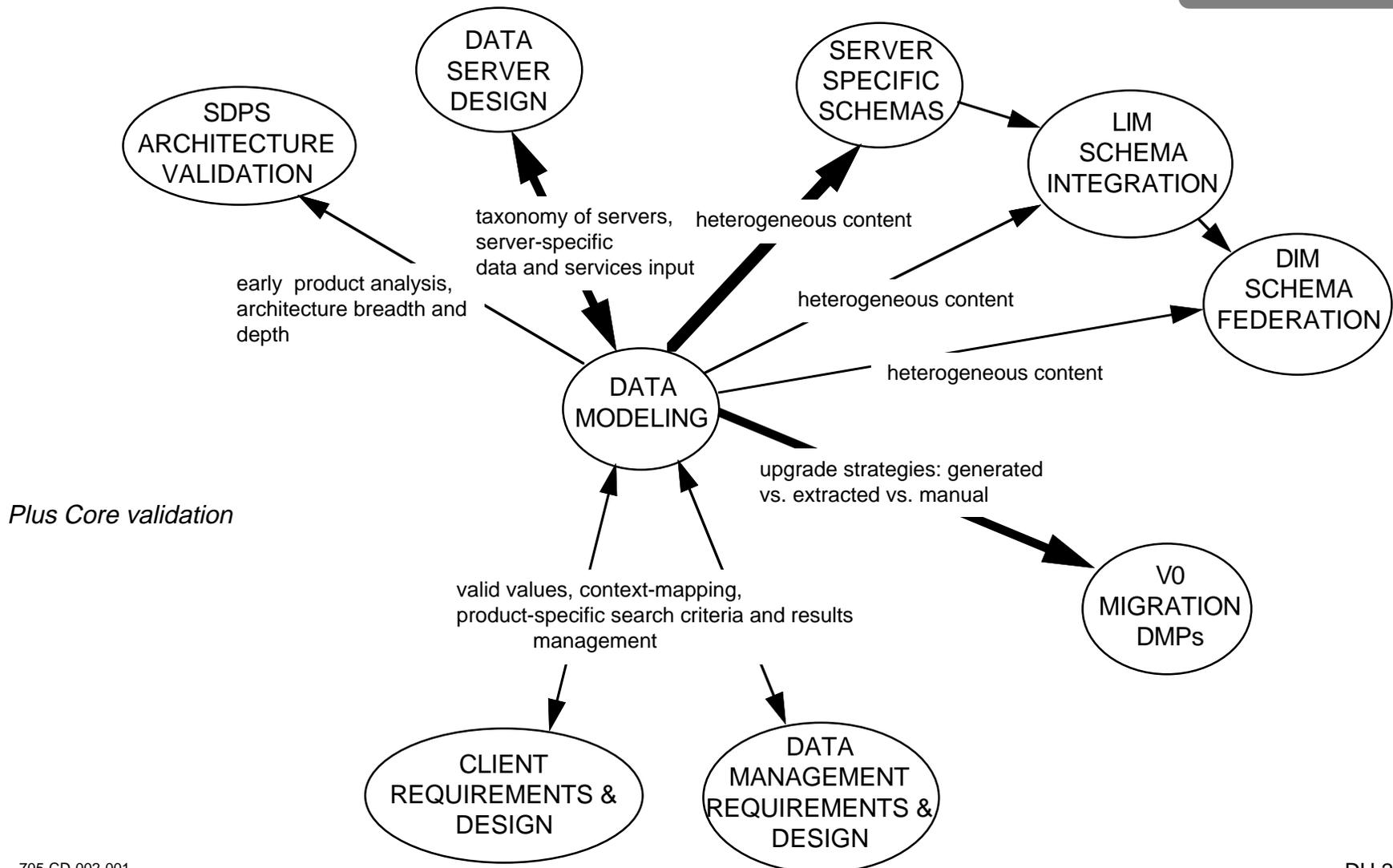
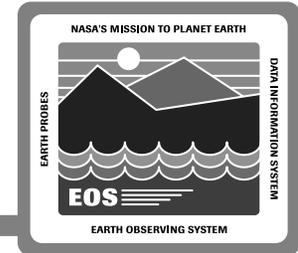
- **Product-specific Context**
- **Development of Data Server Taxonomy**
- **Collection Integration**
- **Current Status**
- **Version 0 Migration Support**

Document References:

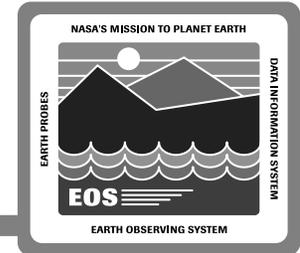
-420-WK-001-001, <http://edhs1.gsfc.nasa.gov> "SDPS Database Design and Database Schema Specification", Working, 1/94 (will become 311-CD-002 as of 3/22/95) Appendix B

-420-TD-001-002, <http://edhs1.gsfc.nasa.gov> "Logical Data Server Taxonomy for PDR"

Product-Specific Context

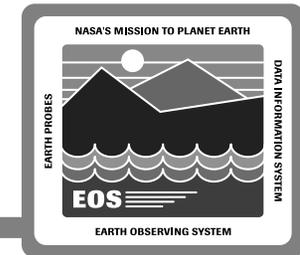


Data Server Taxonomy



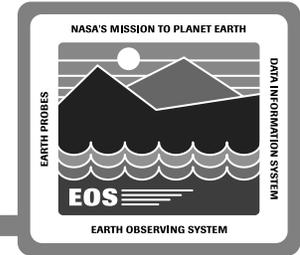
- **Objectives**
 - Ease of addition (e.g. from SCFs)
 - Access pattern support
- **Organization criteria**
 - Instrument + Discipline + Site
 - Attribute/Service similarities
- **Presented at SDR DOAFT, again at DMWG telecon 12/20, 1/6**
 - Need higher-level disciplines
 - Group heritage datasets with EOS timeframe follow-ons
 - Eliminate Version 0 / Ancillary overlaps
- **Process/Impact of updating allocation post-implementation**
- **Document references:**
 - 420-TD-001, "Logical Data Server Taxonomy for PDR"
 - 194-WP-908-001 (MR9402V1), "Data Product Analysis, Early Results", 4/94
 - 194-00293, "Collection Redefinition Impact Study"

Sample Collection: Landsat



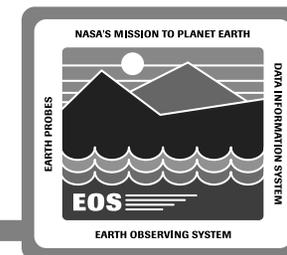
Server	Archive Site	Instrument	Discipline	AHWGP ID	SPSO ID	Product Name	Platform
30--EDC, Landsat, Land Sfc. Veg	EDC	ETM	Land-Surface Vegetation		L0R	--	Landsat-7
	EDC	TM	Land-Surface Vegetation		E-1	NASA Global Change Landsat Data Collection	Landsat 4 and 5
	EDC	TM	Land-Surface Vegetation		E-10	Global Land Cover Test Sites	Landsat 4 and 5
	EDC	TM	Land-Surface Vegetation		E-7	Level-2 "Consistent P" Products	Landsat 4 and 5
	EDC	TM	Land-Surface Vegetation		E-8	Humid Tropical Forest Products	Landsat 4 and 5
	EDC	MSS	Land-Surface Vegetation		E-9	North American Landscape Characterization	Landsat 4 and 5

Collection-Specific Metadata: Example GSFC-TRMM Atmos Dyn



- **Contact Module**
 - 3 aliases, new Valid Value
 - example: TRMM Team Leader to domain of Contact Job Position
- **Data Objects Module**
 - 9 aliases, 3 new attributes, 7 valid values
- **Delivered Algorithm Module**
 - 3 aliases, 2 new attributes
 - example: Calibration Quality attribute added to Calibration Files, possible new core attribute
- **Locality Spatial Module**
 - 5 aliases, 4 new attributes
- **Locality Temporal Module**
 - 11 aliases
- **Source Module**
 - 7 aliases, 2 new attributes, 1 new object class with 5 attributes and their domains
 - example: Keplerian Elements, Guidance Flags, Solar Beta Angles

Collection-Specific Metadata: Example GSFC-TRMM Atmos Dyn



B.1 GSFC-TRMM Collection

This section includes the server specific characteristics, by module, which were applied to the core metadata models for the GSFC-TRMM servers.

The sources of input for the above collection was the population worksheets which were provided by MSFC coupled with server specific analysis which has been an ongoing activity on the EOS Project. The remaining subsections, address by module, the extensions to the core model which were included in the data model for the GSFC-TRMM servers. These extensions included identification of aliases for the core attribute names, several additional domain of values, new attributes and object classes .

B.1.1 Contact Module

The following aliases were added to the core model for this collection

- alias 'Name' added to Contact Name
- alias 'Organization' added to Contact Organization Name
- alias 'Responsible Team' and 'Investigator' added to Role

The following Domain of Value was added to the core model for this collection

- domain of value 'TRMM Team Leader' added to Contact Job Position

B.1.2 Data Objects Module

The following aliases were added to the core model for this collection

- alias 'Granule ID' added to ID of ECS Data Granule
- alias 'Product Type ID' added to Geophysical Parameter Keyword, ID of ECS Data Object, and Processing Level Description
- alias 'Processing Level' added to Processing Level ID
- alias 'Quality Certification Comment' added to Science Review Status
- alias 'Granule Size' added to Size MB of ECS Data Object
- alias 'Associated Datasets' added to ID of ECS Data Granule
- alias 'Dataset ID' added to ID of ECS Data Collection
- alias 'Dataset Name' added to Name of ECS Data Object
- alias 'Product Name' added to Name of ECS Data Object
- alias 'Product Sequence Number' added to ID of ECS Data Object
- alias 'Resolution' added to Direct Spatial Ref Method

The following new attributes were added

- attribute 'Processing Location' added to Processing Level object
- attribute 'Version ID' added to ECS Data Collection
- attribute 'Generation Date' with alias 'Inventory Date' added to ECS Date Object

The following domain of values were added to the core model for this collection

- domain of value 'Day_Night' and 'Cloud_Base' added to Non-Core Attribute Name
- domain of values: 'Missing Data', 'Number of Data Gaps', 'Quality Flags', 'Percentage of Missing Scans', and 'Percentage of Bad or Missing Granules' added to QA Parameter Name

The following new object class with associated attribute and alias was added

- object 'Media' with attribute 'Type' and alias 'Storage Medium' added -- connecting to ECS Data Object

The following constraints were added

- constraint 'Must have QA Parameter Name' added to QA Parameter Value

B.1.3 Delivered Algorithm Module

The following aliases were added

- alias 'Description' added to Algorithm Package Description
- alias 'Algorithm Version' added to Algorithm Package Version
- alias 'Algorithm ID' added to Algorithm Package Name
- alias 'Algorithm Version' added to Algorithm Package Version

The following attributes were added to the core model for this collection

- attributes 'Input Files', 'Date of Generation of Input Files', 'Data Center Source of Input Files', and 'Data Type' added to Input object
- attribute 'Calibration Quality' added to object Calibration Files

B.1.4 Locality Spatial Module

The following attribute was added

- alias 'Spatial Coverage' added to Bounding Coordinate Name

The following domain of values were added

- domain of values 'Min Lon', 'Min Lat', 'Max Lon', 'Max Lat' added to Bounding Coordinate Name

The following new attributes were added

- attributes 'Start Orbit Number', 'Stop Orbit Number', and 'Orbit ID' with alias 'Orbit Number' added to object Orbit Calculated Spatial Domain

B.1.5 Locality Temporal Module

The following attributes were added to the core model for this collection

- alias 'Temporal Coverage', 'Coverage Dates', 'Start Date' added to Beginning Date
- alias 'Temporal Coverage', 'Start Time' added to Beginning Time
- alias 'Temporal Coverage', 'Coverage Dates', 'Stop Date' added to Ending Date
- alias 'Temporal Coverage', 'Stop Time' added to Ending Time
- alias 'Temporal Coverage' added to Time of Day
- alias 'Recording Date' and 'Processing Date' added to Calendar Date

B.1.6 Reference Paper Module

No Changes

B.1.7 Source Module

The following aliases were added

- alias 'Sensor ID' added to Source Short Name
- alias 'Platform ID' added to Satellite Short Name
- alias 'Minimum Valid Value of Channel' added to Channel Spectrum Start
- alias 'Maximum Valid Value of Channel' added to Channel Spectrum Stop
- alias 'Source' added to Source Short Name
- alias 'Instruments Center Wavelengths' added to Channel Center Wavelength
- alias 'Number of Scans per Orbit' added to Instrument Operational Modes
- alias 'Source' added to Source Short Name

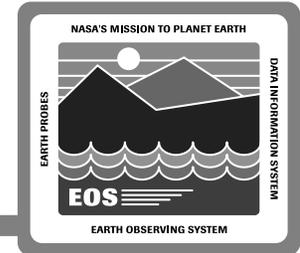
The following attributes were added to existing objects in the core model for this collection

- attribute 'Source ID' with alias 'Project ID' added to Source object
- attribute 'Channel Data Accuracy' with alias 'Accuracy' added to Channel object

The following object classes and associated attributes with domain of values were added

- object 'Satellite Position' added which connects to Satellite object and has the following attributes: 'Keplerian Elements', 'Guidance Flags', 'Solar Beta Angles', 'Scanner Antenna Beam Alignment and Calibration Parameters', 'Attitude Information'
- attribute Keplerian Elements has the following domain of values: 'Epoch Time', 'Semi-Major Axis', 'Eccentricity', 'Inclination', 'Right Ascension of Ascending Node', 'Mean Anomaly', 'Argument of Perigee', 'Mean Motion'
- attribute Guidance Flags has the following domain of values: 'Orbit Adjust Flag' and 'Attitude Mode Flag'

Prod-Spec Metadata Status



- **GSFC Servers:**

- GSFC-TRMM Collection
- GSFC-VIRS L0/L1 Collection
- GSFC-CZCS L0/L1 Collection
- GSFC-TOMS-ATMOS_COMP Collection

- **MSFC Servers:**

- MSFC SMMR L0/L1 Collection
- MSFC_TSDIS Collection
- MSFC_PR L0/L1 Collection
- MSFC_TMI L0/L1 Collection
- MSFC_RADAR TAPES, L0/L1 Collection
- MSFC_V0 Collection

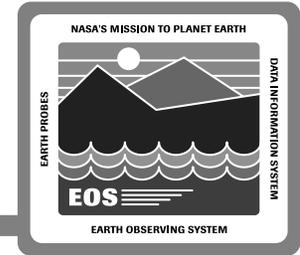
- **LaRC Servers:**

- LaRC CERES ERP Collection
- LaRC CERES L0/L1 Collection
- LaRC ERBE ERP Collection
- LaRC ISCCP Collection
- LaRC SAGEII(87) Collection
- LaRC SAGEII(88) Collection
- LaRC GTE ABLE Collection
- LaRC FIRE Collection

- **EDC Servers:**

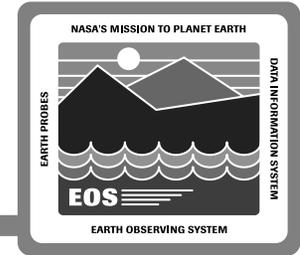
- EDC Landsat Land Sfc Veg Collection

Version 0 Migration Support



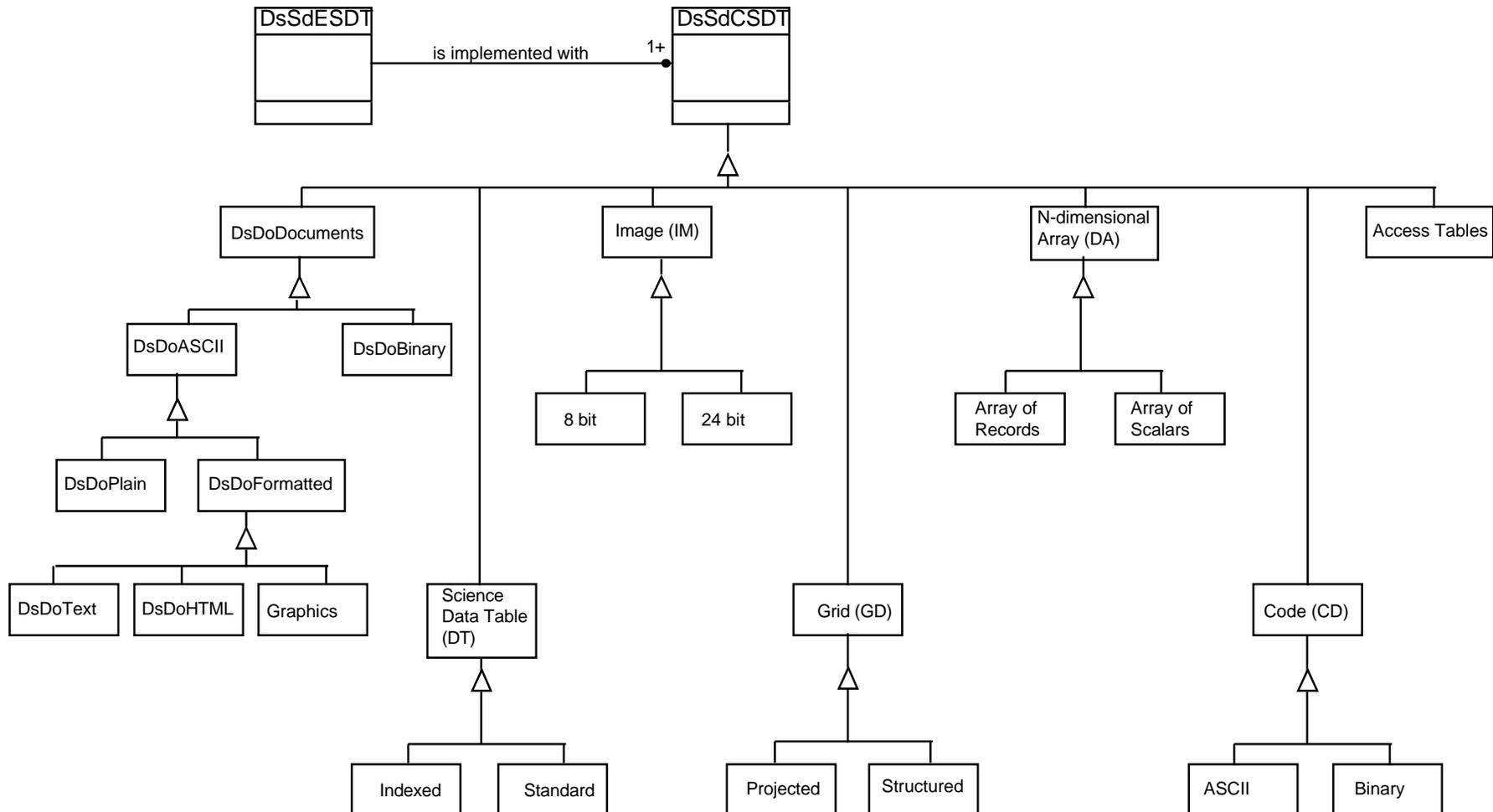
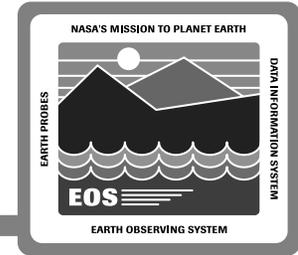
- **Metadata Worksheets, DAAC Scientists**
- **Aliases Identified**
- **Product/Collection Specific Metadata found**
- **Core metadata validated**
- **Input to Cost Model: Generation mechanism**
- **Input to Data Migration Plans (DMPs): Extract vs. Generate vs. Manual definition**

Data Formats Agenda

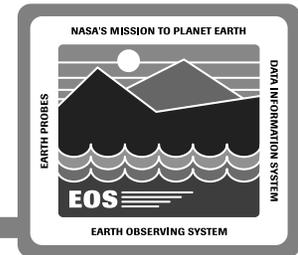


- **CSDT Data Server support, Release A**
- **Product Data Types Strawman**
- **EOS View Data Type support, Release A**

Data Formats: CSDTs



Data Formats: Data Type Matrix



Product (ESDT)	TRMM	CER01	CER02	CER03
Major CSDT	native	swath	swath	grid
Name	CERES Level 0 data	Bi-Directional Scan (BDS)	ES-8 ERBE-Like Product	ES-9 ERBE-Like Product
Level	L0	1B	2	3

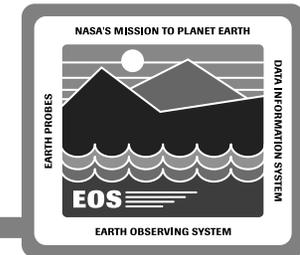
• • •

•
•
•

Document Reference:

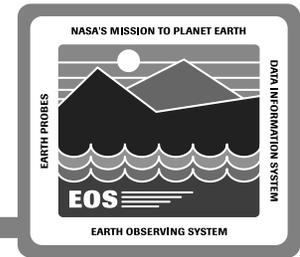
- 304-CD-002-001, <http://edhs1.gsfc.nasa.gov> "SDPS Level 4 Requirements", 1/94, Appendix F

Data Formats: EOSView



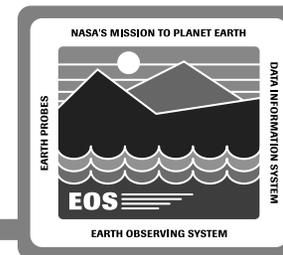
- Core Technologies
 - + Scripting
 - + Interface
 - Other programs pass commands to EOSView
 - EOSView passes information back to other programs
 - + Script commands
 - Open, Close, Animate
 - List Contents, Display selected object, display first object
 - Add Overlay, report clicked location, change colortable, etc.
 - + Pseudocolor Display
 - + Color Palette
 - Fiddle color palette
 - + **Overlays** (IMS0580, IMS1530)
 - + Overlay Lists
 - Store Continents Lat/Long outlines
 - + Vector Overlay Display
 - Display Continent outline as vectors on grid objects
 - + Lat/Long Grid Display
 - Display Lat/long gridlines on grid objects
- Data Display Windows
 - + **Metadata** Display Window
 - Display metadata in scrolling window
 - Display both text and HDF attribute metadata
 - + **Table** Display Window (IMS0690)
 - + Features
 - Display column names, row numbers in scrolling spreadsheet list
 - + Supported Structures
 - Vdatas, Science Data Tables, Point Datas
 - + **Text** Display Window (IMS0690)
 - + Features
 - Display text in scrolling window
 - + Supported Structures
 - text, data dictionary
- Visualization Windows
 - + **Point** display
 - + Components
 - Display point structure components in a list
 - Select a component and display appropriately
 - Possible: text, metadata, tables
 - + 2D Color Scatter (IMS1540)
 - Display point data as colored symbols on grid
 - + **Grid** display
 - + Components
 - Display grid components in list
 - Select component and display appropriately
 - Components: text, metadata, arrays
 - + Features
 - Display simple grid as pseudocolor
 - Display complex grid as pseudocolor
 - Display overlay on grid
 - + Lat/long Conversions
 - Convert from Lat/long to pixel location for overlays
 - + **Swath** display
 - + Components
 - Display swath components in list
 - Select component and display appropriately
 - Components: text, metadata, arrays
 - + Features
 - Display array as pseudocolor
 - (no overlays)
 - + **Animation**
 - Animate from disk or from memory
 - Specify an animation file or a user created animation directory
 - Enable palette manipulations (for all images)

Next Steps



- **Continued Product-Specific Analysis and Design**
- **System Administration Data Design**
- **Database Physical Design and Implementation**
- **Metadata Population Control Authority**

System Administration Data Design



Resource Model of 420-WK-001

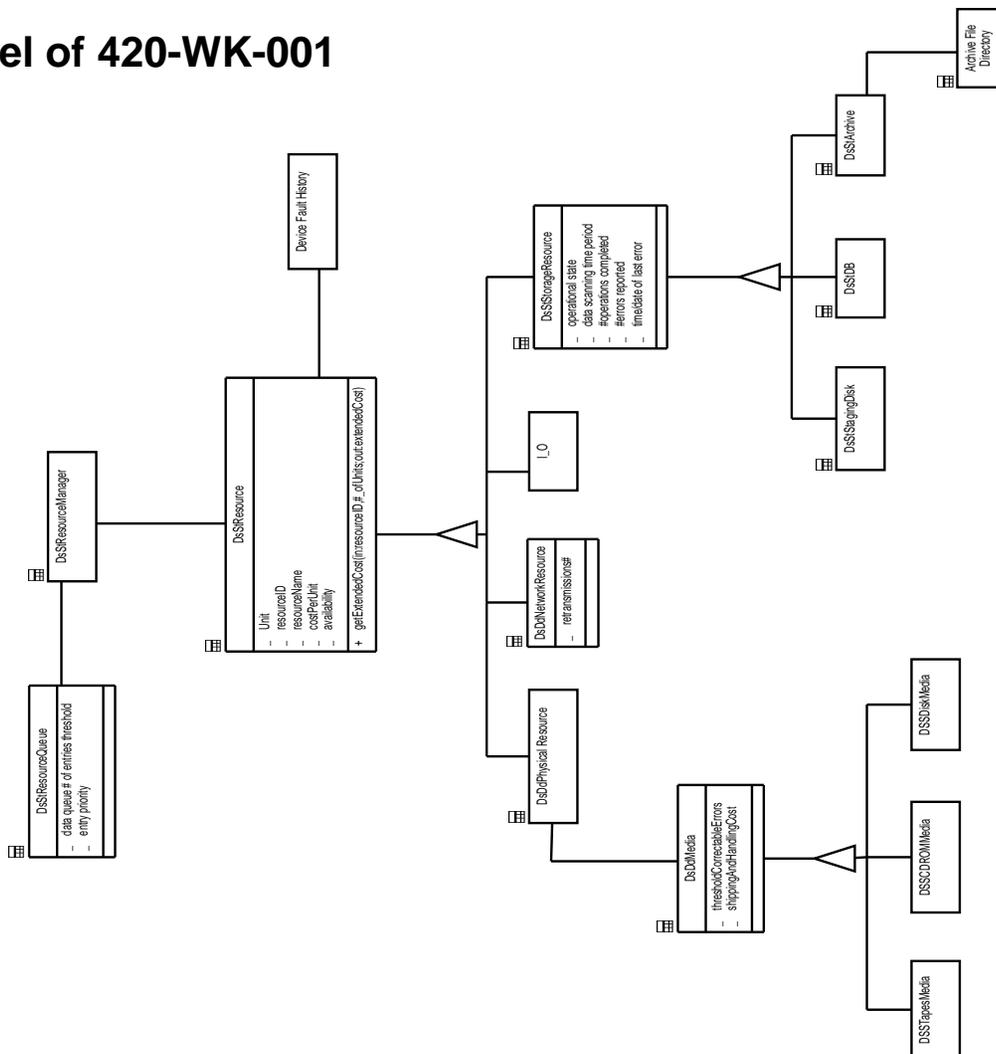
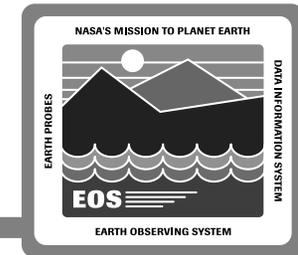


Figure 5-2. SDPS Resources

Data Modeling Status



Data Category	Supported by ECS at Release:	Analysis Timeframe
"Release A" Version 0	At initial Release A	PDR, 3/95
TRMM Datasets (CERES, LIS, PR, TMI, VIRS)	At initial Release A	PDR, 3/95
TRMM Ancillary data	At initial Release A	PDR, 3/95 as known
Landsat 7 LOR product	At Initial Release B	PDR, 3/95
High Priority Version 0	Between Release A and B	PDR, 3/95; CDR 8/95
AM1 Platform At-Launch Products	At initial Release B	PDR, 3/95 as known; Release B IDR
AM1 Platform At-Launch Ancillary Data	At initial Release B	PDR, 3/95 as known; Release B IDR
All other Version 0	Between Release B and EOY1999	Release B IDR
SeaWiFS II	Between Release B and EOY1999	Release B IDR
AM1 Post-Launch Products	Post Release B	Release B IDR
AM1 Post-Launch Ancillary data	Post Release B	Release B IDR
Remaining Products	IDS products start with Release A; others include PM platform, AM2, CHEM, ACRIMSAT, LASER ALT, et al	post Rel. B IDR; see Data Tracking Matrix for Detail

