



National Aeronautics and Space
Administration
Jet Propulsion Laboratory
California Institute of Technology



WISE Science Data Center CDR

WSDC Overview, Requirements and Implementation

R. Cutri - IPAC

WISE Science Data Manager

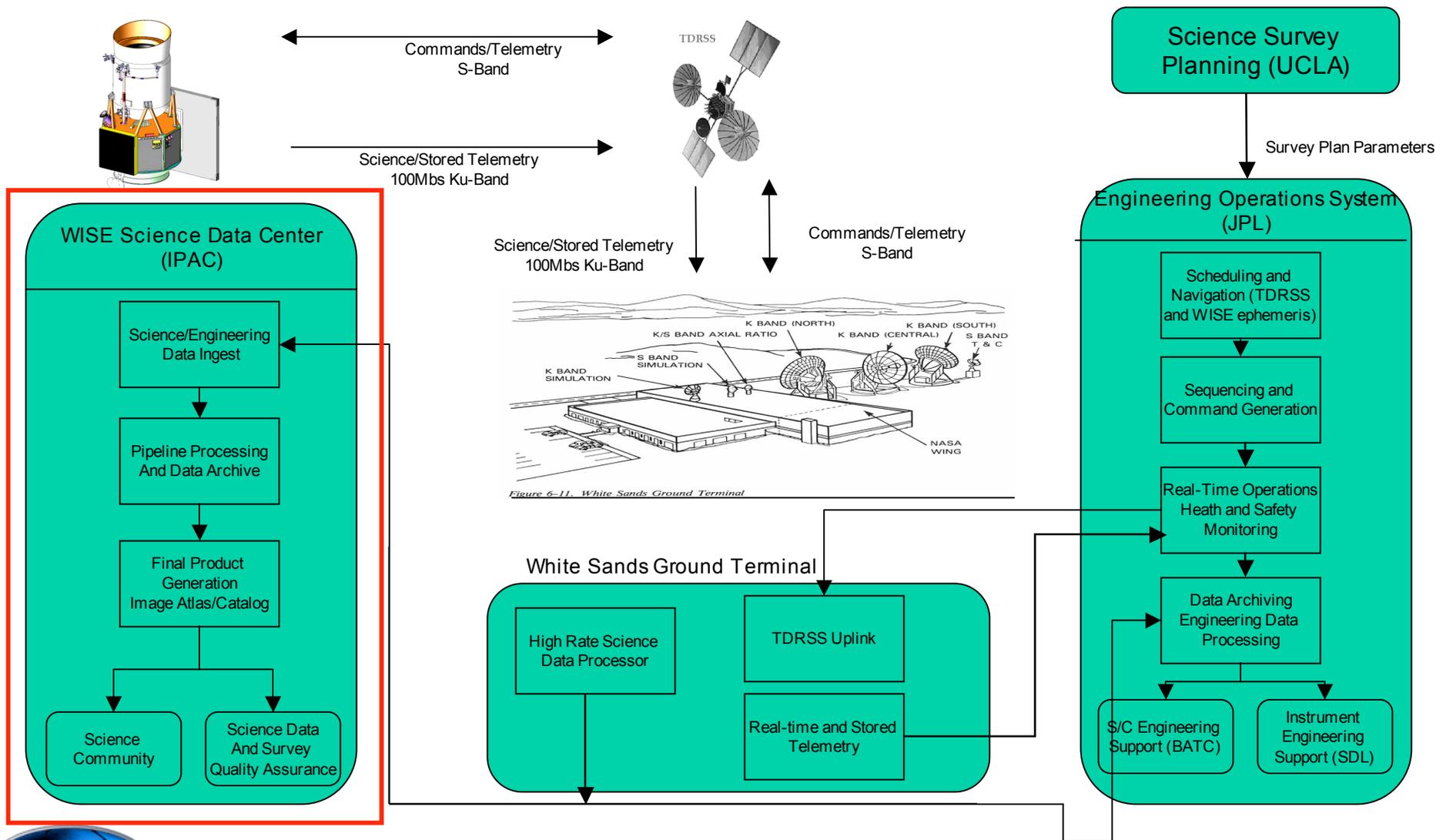




MOS Architecture



WSDC Overview





WSDC Responsibilities



- **Data Ingest**
 - Receive raw science image packets from White Sands and engineering telemetry from MOS
 - Decompress and assemble image data into Level 0 FITS format, and mate with engineering telemetry
- **Science Data Processing**
 - Convert raw imaging and engineering data into a photometrically and astrometrically calibrated Image Atlas and extracted Source Catalog
 - Compile Explanatory Supplement to the WISE Data Products
 - Generate ancillary data products to support mission requirements
- **Science Data Quality Assurance**
 - Provide rapid QA feedback for selected on-orbit performance (i.e. scan mirror synch)
 - Overall science data QA for survey planning and data product generation
- **Science Data Archiving and Distribution**
 - Archive raw data (Level 0) during mission
 - Provide an operational archive to store and serve intermediate data products and metadata to project team during mission
 - Provide a long-term “living” archive that stores and serves WISE science data products to the astronomical community and general public along with user’s guide documentation and descriptive analysis





WSDC Overview

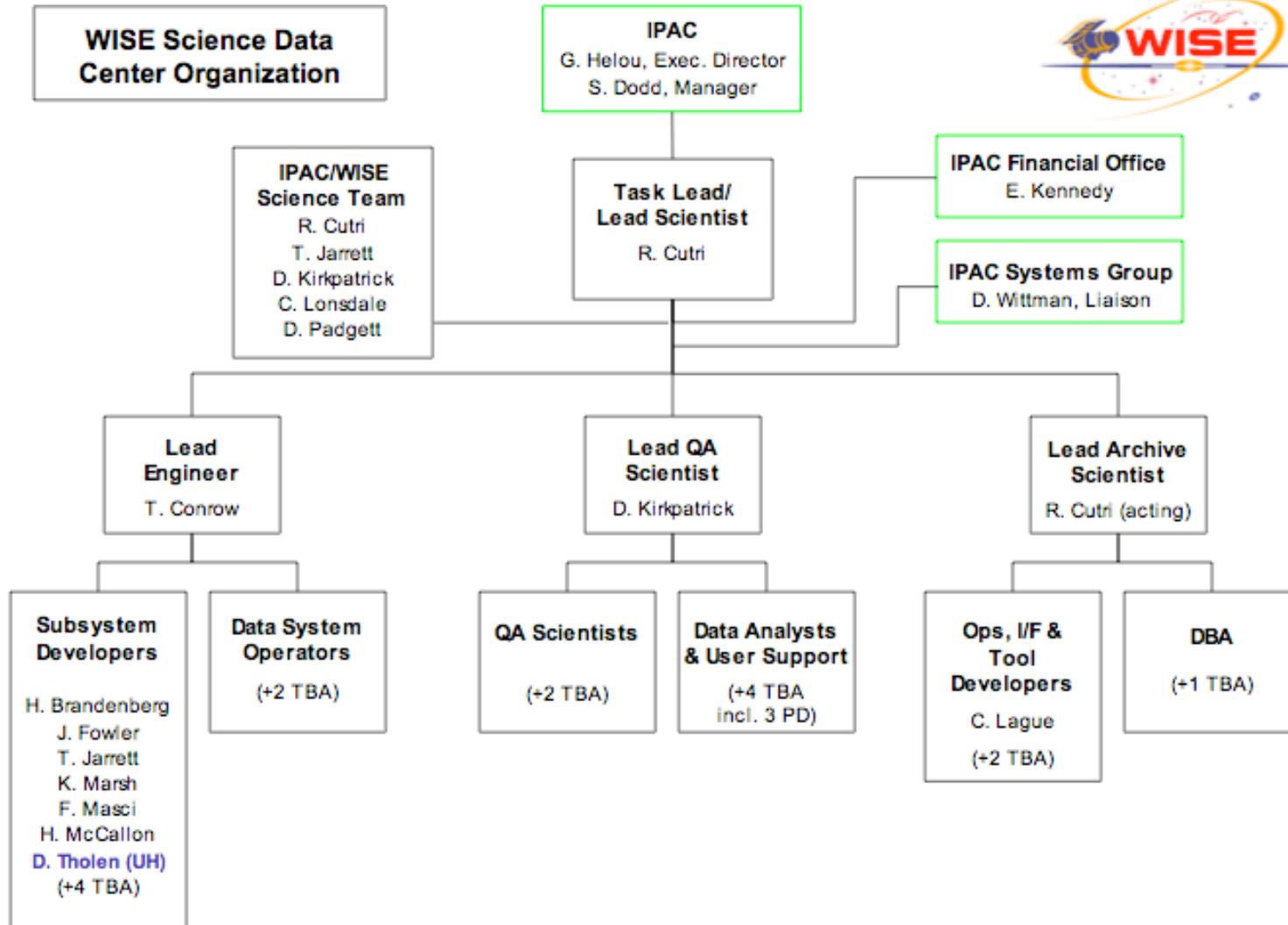


- The WSDC is an independent *task* at IPAC
 - Analogous to 2MASS, NHSC, IRSA, NED
 - IPAC/WISE Task Lead reports to IPAC Executive Director, George Helou, and works with IPAC Manager, Suzanne Dodd, to organize staff and resources to perform task
 - Administrative, facilities and common infrastructure cost shared with all IPAC projects
 - WSDC will use dedicated hardware that is integrated into IPAC network structure
 - Resources and expertise shared with other IPAC tasks along with the Spitzer Science Center and Michelson Science Center which reside under a common administrative umbrella at the “Greater IPAC”





WSDC Organization





WSDC Status



- WSDC staffing and activity was kept at minimal level until start of Phase C/D
 - Phase B work focused on cost and schedule refinement, requirements definition, high-level system design and key data and operational interface definition (MOS-GDS/WSDC)
- Phase C/D WSDC staff ramp-up is underway
 - 7 of projected 12 development staff on-board. 4 to be added in FY08.
- Requirements defined and traced to higher levels
- Mature high-level system design
- Mature designs and prototypes completed for many subsystems/modules
 - Peer reviews complete for subsystems with significant new features
 - Selected additional peer reviews to follow CDR
- Work on some subsystems intentionally scheduled for start later this year (e.g. ARCHIVE, FPG, artifact ID, photometric calibration)
 - Conceptual designs for some presented in this CDR (ARCHIVE, WPHOT, PCAL)
 - No separate presentations will be made for FPG, ARTID (these elements are in development plan, and resource are allocated)





Key Activities Leading up to WSDC CDR



- Successful Mission CDR (6/07), MOS CDR (7/07)
- Finalized WSDC Functional Requirements (FRD v2)
- Advanced versions of all top-level documents (PDMP, FDD, SMP, QA Plan)
- Project-level Calibration Peer Review (9/07)
 - Included ICAL, PCAL conceptual designs
- WSDC peer reviews of PIPELINE subsystems with significantly new algorithms components
 - AWAIC (11/07), MDET (12/07), PREX (12/07)
- Demonstrated HRP science data INGEST in support of Spacecraft Mission Unique Board (MUB) testing - primary external interface (2/07)
- WSDS v0 delivered (10/07)
 - Internal data flow, operational data structure demonstration
 - Selected subsystem prototype functionality (detection, position reconstruction)
- Data product formats (image specifications/headers, catalog columns) discussion started with Science Team
- Support of on-going detector testing to assist with flight FPA selection





Changes Since MOS PDR



- Launch date moved to November 2009
- Extended Phase B completed
- Phase C/D started February 2007
- Added prototype ingest system development to support MUB testing
- Ground test data archive deployed at WSDC
- Profile-fit photometry introduced into baseline processing design





Driving Requirements - 1

excerpted from WSDC FRD (WSDC D-R001)



ID	Higher Level Requirement	Level 4 Requirement	Verification Method
L4WSDC-001	L3MOS-366 (L1PP-8)	Image Atlas: The WSDC shall produce a digital Image Atlas that combines multiple survey exposures at each position on the sky.	Demonstration
L4WSDC-002	L3MOS-374 (L1PP-9)	Source Catalog: The WSDC shall produce a Source Catalog derived from the WISE digital Image Atlas.	Demonstration
L4WSDC-003	L3MOS-366,374 (L1PP-34)	Data Release: The WSDC shall release to the public an image atlas and source catalog covering the full survey area within 17 months after the end of on-orbit data collection.	Demonstration
L4WSDC-004	L3MOS-355, 363 (L1.5SRD-50)	Preliminary Data Release: The WSDC shall release to the public a preliminary image atlas and source catalog covering at least 50% of the surveyed area within 6 months after the end of on-orbit data collection	Demonstration
L4WSDC-080	L3MOS-417 (L1PP-010)	Catalog Reliability: The final WISE Source Catalog shall have greater than 99.9% reliability for sources detected in at least one band with SNR>20, where the noise includes flux errors due to zodiacal foreground emission, instrumental effects, source photon statistics, and neighboring sources. This requirement shall not apply to sources that are superimposed on an identified artifact.	Demonstration: Comparison with external "truth tables"; detection confirmation statistics
L4WSDC-009	L3MOS-363, 418 (L1PP-011)	Catalog Completeness: The final WISE Source Catalog shall be at 95% complete for sources detected with SNR>20 in at least one band, where the noise includes flux errors due to zodiacal foreground emission, instrumental effects, source photon statistics, and neighboring sources. This requirement shall not apply to sources that are superimposed on an identified artifact.	Demonstration: Detection repeatability in deep coverage areas (e.g. the ecliptic poles)



Driving Requirements - 2

ID	Higher Level Requirement	Level 4 Requirement	Verification Method
L4WSDC-013	L3MOS-376 (L1PP-012)	Photometric Accuracy: The root mean square error in relative photometric accuracy in the WISE Source Catalog shall be better than 7% in each band for unsaturated point sources with SNR>100, where the noise flux errors due to zodiacal foreground emission, instrumental effects, source photon statistics, and neighboring sources. This requirement shall not apply to sources that superimposed on an identified artifact.	Demonstration: Photometric repeatability; stellar color stability over sky
L4WSDC-014	L3MOS-370 (L1PP-013)	Astrometric Accuracy: The root mean square (1σ) error in WISE catalog positions with respect to 2MASS All-Sky Point Source Catalog positions shall be less than 0.5" on each axis	Demonstration: Comparison with 2MASS PSC and other astrometric catalogs (e.g. UCAC)
L4WSDC-012	L3MOS-420 (L1PP-4)	Photometric Sensitivity: Flux measurements in the WISE Source Catalog shall have a SNR of five or more for point sources with fluxes of 0.12, 0.16, 0.65 and 2.6 mJy at 3.3, 4.7, 12 and 23 micrometers, respectively, assuming 8 independent exposures and where the noise flux errors due to zodiacal foreground emission, instrumental effects, source photon statistics, and neighboring sources.	Demonstration: Photometric repeatability; Colors of normal stars, galaxies
L4WSDC-065	L3MOS-272	Quicklook Quality Assurance: A sample of 3% of the imaging data returned to the ground each day shall be processed in an expedited way to produce a Quicklook report that monitors the routine performance of the flight system as can be determined from the science data, and identifies problems that may require prompt action by WISE Science or Mission Operations.	Demonstration



Driving Requirements - 3



ID	Higher Level Requirement	Level 4 Requirement	Verification Method
L4WSDC-039	L3MOS-284	Single Orbit Data Processing Latency: Within 3 days from receipt of a given data set at the WSDC all data be shall processed through the WSDS Scan/Frame pipeline which performs basic image calibration and source extraction on images from individual orbits. The results of this processing step shall be Level 1 source extracted source Working Database (L1WDB) and Image Archive allowing access by the WISE Science Team for external quality assessment.	Demonstration
L4WSDC-053	L3MOS-387 (L1PP-34)	Archive Longevity: The WSDC shall make the Image Atlas and Catalog products accessible to the astronomical community in collaboration with the NASA/IPAC Infrared Science Archive (IRSA) to ensure long-term availability beyond the end WISE missions operations, and to ensure interoperability with other NASA mission archives.	Demonstration





Key Deliverables - 1



- WISE Image Atlas
 - Formed by combining all image frames covering each point on the sky
 - FITS format, 4 bands registered, approx. $1^\circ \times 1^\circ @ 1.375''/\text{pix}$.
 - Metadata describing each Atlas Image
 - Approx. 140,000 images covering entire sky in four WISE bands (depends on final footprint)
- WISE Source Catalog
 - Basic attributes of each object detected on combined (*i.e.* Atlas) images
 - Position on the sky (J2000) and associated uncertainties
 - Photometry in four WISE bands and associated uncertainties, flux upper limits in non-detected bands
 - Source detection and measurement quality flags and parameters (*e.g.* detection statistics, reliability estimate, photometric quality, confusion and contamination, shape info)
 - A unique identifier (*i.e.* source name)
 - Additional information to enhance usability (*e.g.* association with 2MASS)
 - Current estimate is ~300 million entries (*Working Database* will contain $\sim 10^9$ entries)





Key Deliverables - 2



- WISE Explanatory Supplement
 - Mission and data product description,
 - User's guide (*e.g.* data formats, access modes)
 - Cautionary notes
- Ancillary Products
 - Not specified in higher level requirements. To be developed as resources and schedule allow*
 - Atlas Image coverage and noise maps
 - Solar system object association list (derived from single-epoch images)
- Single-epoch (Level 1) images and extracted source database maintained during mission operational period for Project/Science Team access, *but are not base-lined for public release*





Data Product Delivery



- Two-stage Data Release
 - Preliminary Image Atlas and Source Catalog
 - Derived from first 50% of sky surveyed
 - Conservative Catalog SNR limits
 - Release 6 months after end of on-orbit operations (12/2010)
 - Final Image Atlas and Source Catalog
 - Derived from all survey data
 - Release 17 months after end of on-orbit operations (11/2011)
- Data Access Mode (project team, science community and general public)
 - On-line services of NASA/IPAC Infrared Science Archive (IRSA)
 - IRSA will host Level 1, 3 (for Project Team access) and final product archive (for public access)
 - Option to provide bulk distribution of Source Catalog





Key Receivables

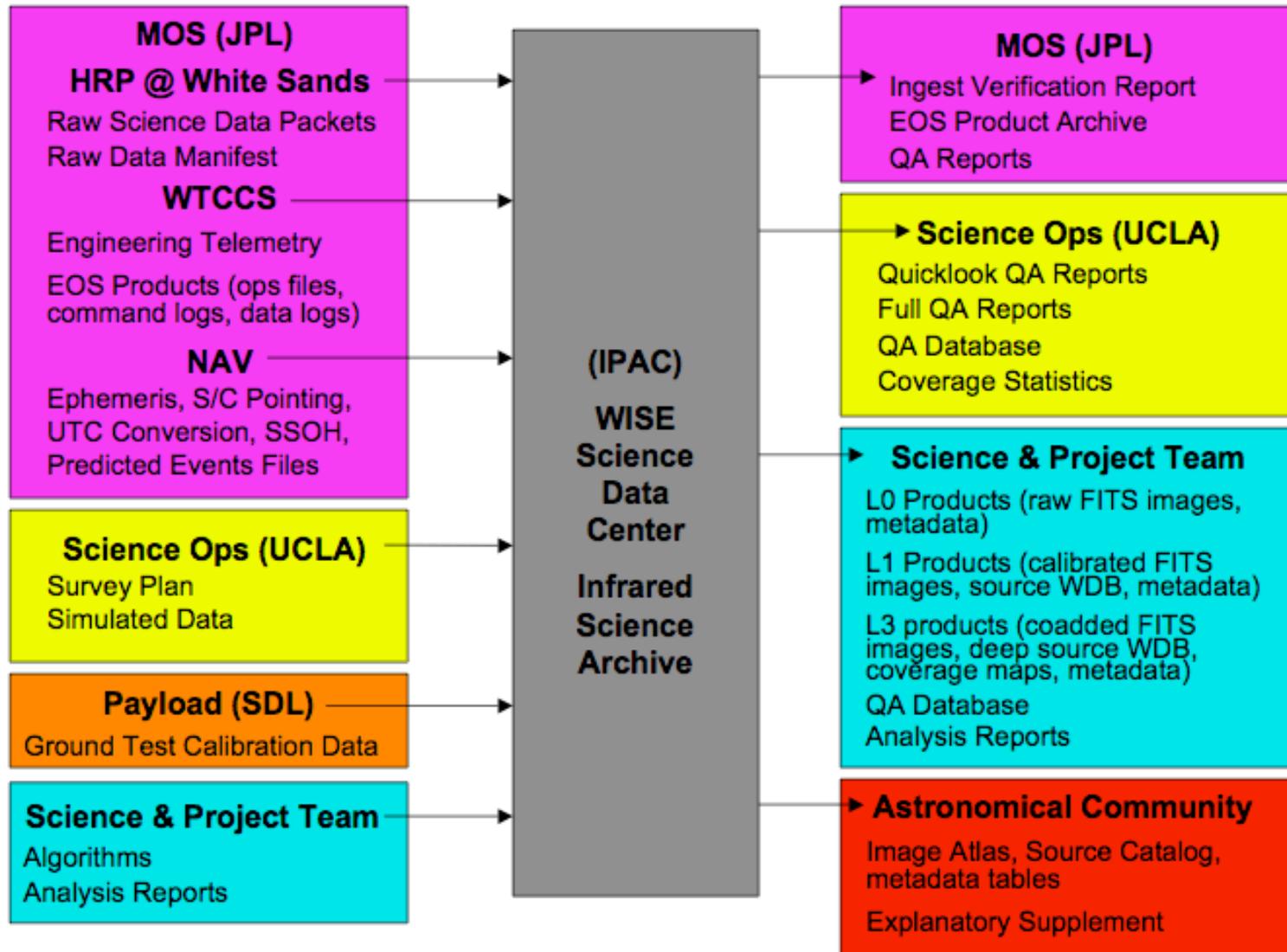


- System interface specifications (MOS)
 - *MOS GDS ICD (JPL D-34372)*
 - Formats, specifications of time/band-tagged HRP pixel data, time-tagged engineering H/K data and ancillary products
- Calibration products (Project, Science Team)
 - *Calibration Plan Document (JPL D-33753)*
 - Pre-launch and IOC-update instrumental calibration (read noise, masks, darks, flats, linearity, droop, distortion maps, PRF maps)
 - Photometric standard star network
- Simulation data (SOC/UCLA)
 - *Simulated Data Needs Document (WSDC D-I001)*
 - Astronomical scenes with instrumental signatures in four WISE bands
 - Fidelity to evolve as instrument testing progresses
 - Culminates in 30 full consecutive orbits



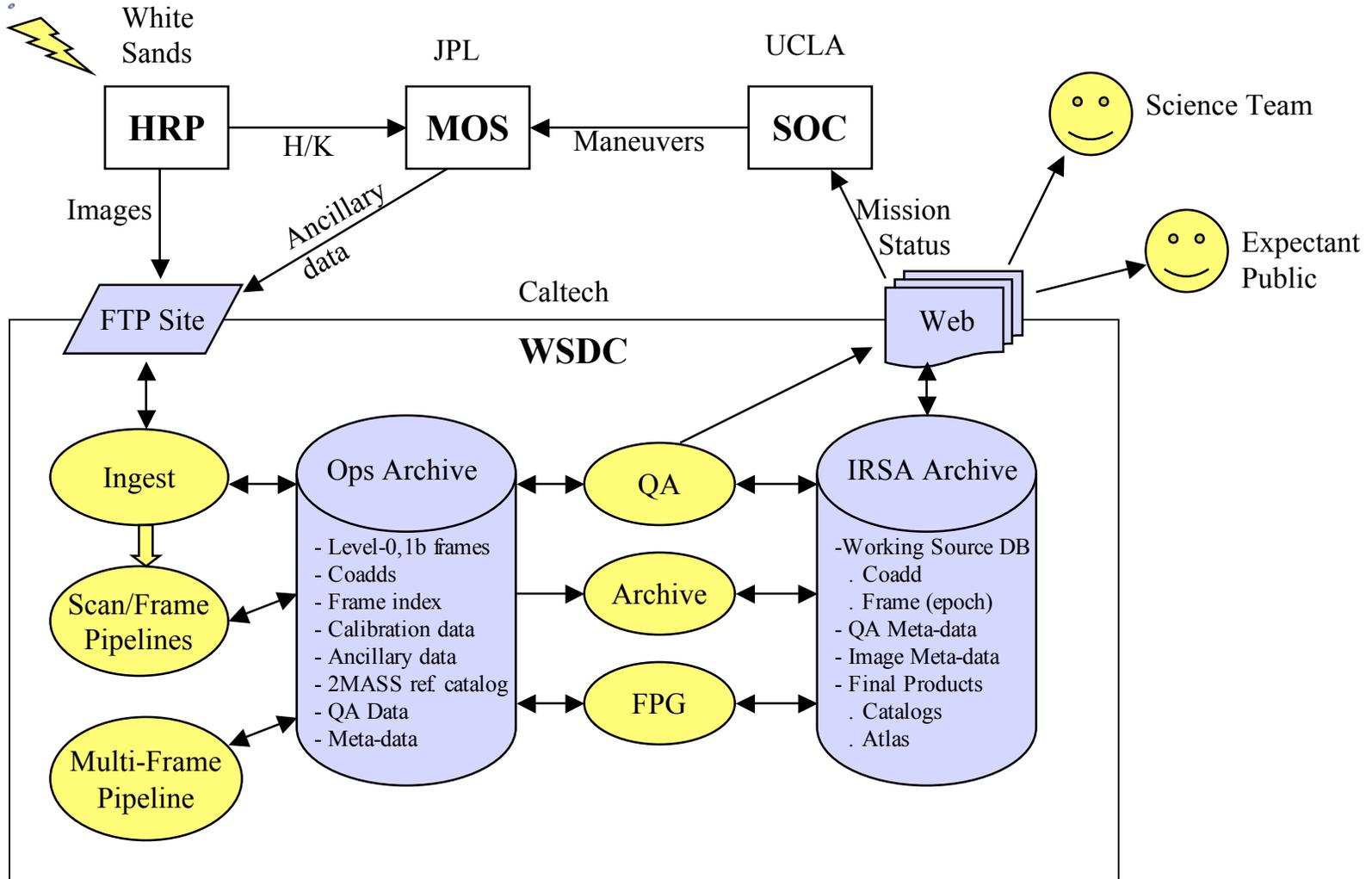


External Interfaces





WISE Science Data System (WSDS)





WSDS Subsystems



- INGEST
 - Receives science data packets and engineering telemetry from MOS and assembles Level 0 FITS-format files. Stages Level 0 images and metadata for pipeline processing.
- Data Reduction PIPELINES
 - Converts Level 0 imaging data into calibrated images and extracted source *Working Databases*
 - **Frame/Scan pipeline** operates on individual frames within one “scan” (=1/2 orbit)
 - **Multiframe pipeline** operates on data from multiple orbits
- Quality Assurance (QA)
 - Generates concise reports summarizing science data quality using summary outputs from other subsystems. Web-based report, with capability to drill-down to detailed image, graphical and tabular data
 - Reports reviewed by QA scientists at WSDC. Final quality assignment approved by PI or designee
- EXEC
 - Provides interface-related services to wrappers and pipelines. Mediates between external callers and applications, providing a uniform interface, binding execution units (modules) together into a unified pipeline
- ARCHIVE/Distribution System
 - Archives raw and processed mission data and metadata. Serves Image Atlas, Source Catalog and mission metadata to WISE project team and astronomical community. Integrated into Infrared Science Archive (IRSA) at IPAC.
- Final Product Generator (FPG)
 - Constructs WISE Preliminary and Final Image Atlases and Source Catalogs from *combined* image and source *Working Databases*. Includes validation, characterization and documentation.





Data Processing Levels

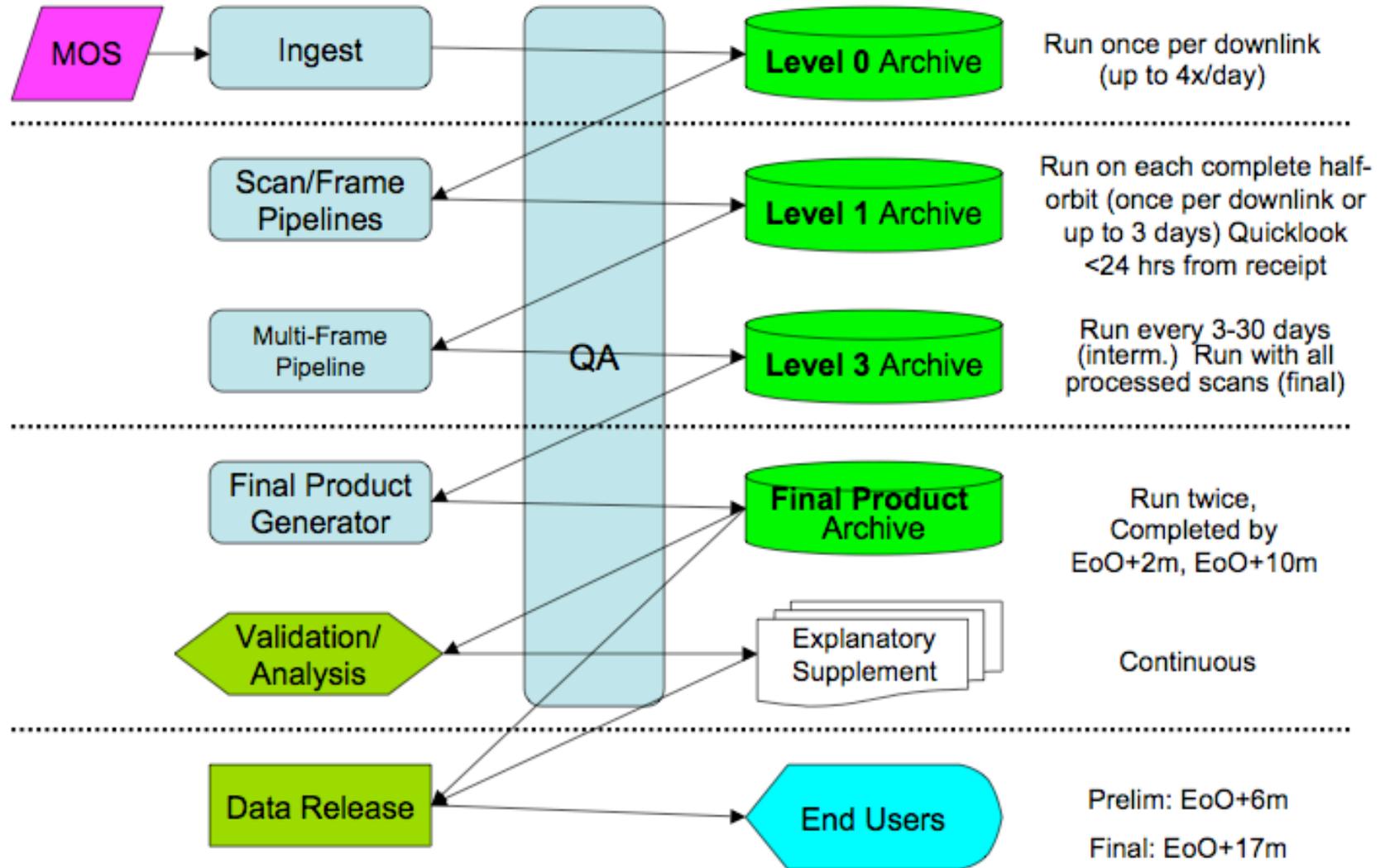


Name	Description
Telemetry	CCSDS source packets
Raw	FITS, integer 2.75" pixels, band+time meta-data
Level-0	Raw + real pixels, add much meta-data
Level-1	Level-0 + instrumental, astrometric/photometric calibration
Level-1a	Level-0 + instrumental calibration applied to pixels
Level-1b	Level-1a + astrometric/photometric calibration in header
Level-2	Level-1 + upsampled and undistorted (rubber-sheeted)
Level-2a	Level-1a + upsampled and undistorted
Level-2b	Level-1b + upsampled and undistorted
Level-3	Multiple frame coadds, upsampled, undistorted
Atlas Images	Selected FPG Level-3 products





Internal Data Flow and Operational Cycle





Key Implementation Features - 1



- Integrated into WISE planning and design from the beginning
 - WSDC participates in project management activities, decisions, strategic planning
- Intellectual heritage from other IPAC-supported missions/projects
 - WISE processing and archiving software will be developed by many of the same engineers and scientists that performed similar tasks on 2MASS, Spitzer, GALEX, IRAS
- Processing, QA, FPG, Archive based closely on systems used for 2MASS
 - Highly automated, “industrial strength” data processing software system designed for high-throughput, reliable operation
 - Extensive use of automated QA reporting
 - Modular system to facilitate parallel development, unit-testing
- Planned two-stage data processing and data release
 - “Can’t get it right the first time”
 - Gets preliminary version of data out to community as rapidly as possible
 - Allows time to incorporate best knowledge of actual instrument performance, calibration and sky for “final” version





Key Implementation Features - 2



- Design from outset with *end product development and distribution* in-mind
 - Leverage IRSA infrastructure to provide easy access to intermediate and final data products for Science and Project Team
 - Interfaces well-tested prior to public releases
- Thoroughly vetted and characterized data products with detailed on-line user documentation (Explanatory Supplement) describing processing algorithms and products
- On-going interaction with WISE Science Team during all phases of the project
 - Five WISE co-Investigators are IPAC staff members (Cutri, Jarrett, Kirkpatrick, Lonsdale, Padgett) that provide strong scientific oversight for WSDC activities
 - “Cognizant” Science Team member assigned to each data processing pipeline subsystem will work closely with WSDC cognizant engineer to develop and validate algorithms, analyze on-orbit data, validate release products (adopted from 2MASS model)
- WSDC software developers remain on staff throughout mission
 - Necessary for two-stage processing strategy - final processing takes place after the end of on-orbit operations
 - Retains personnel with key expertise in WSDC software systems, algorithms and survey data





Key Assumptions



- Ancillary or intermediate data products not specified in the Level 1 or 1.5 requirements will be developed on a best-effort basis (e.g. solar system object association in single-epoch exposures, image coverage maps)
- Software developed at IPAC governed by proven IPAC management methods and standards (*i.e.* not treated as flight software)
- Plan for nominal 7 month mission (1 month IOC + 6 month survey operations), but design must be scalable to 13 month mission, with additional resources (Phase F).
- WSDC activities will continue for 6 months following the final data product release for documentation finalization/clean-up, to provide user-support, and to oversee transfer of data curation to IRSA.

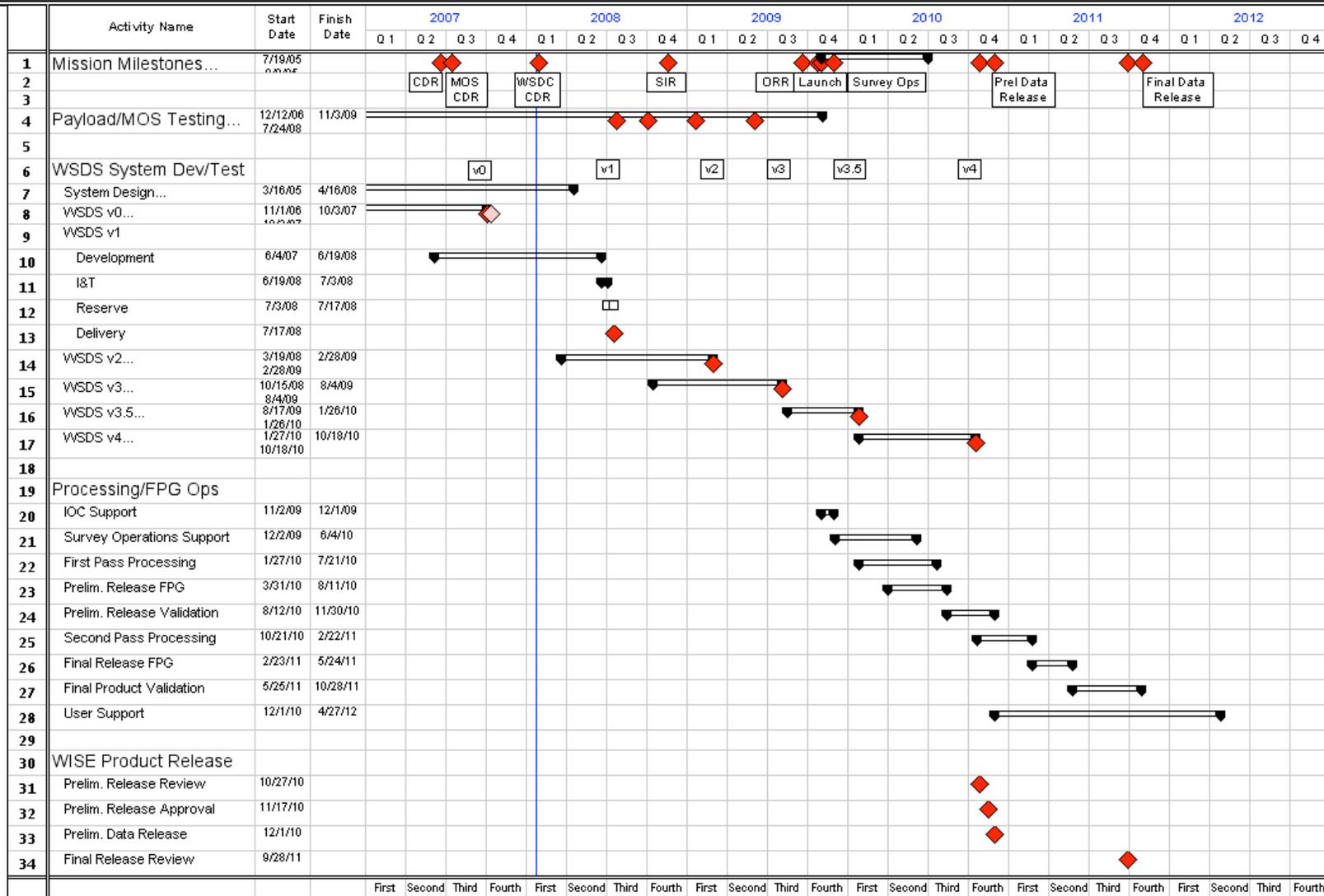




WSDC Top Level Schedule



WSDC Overview

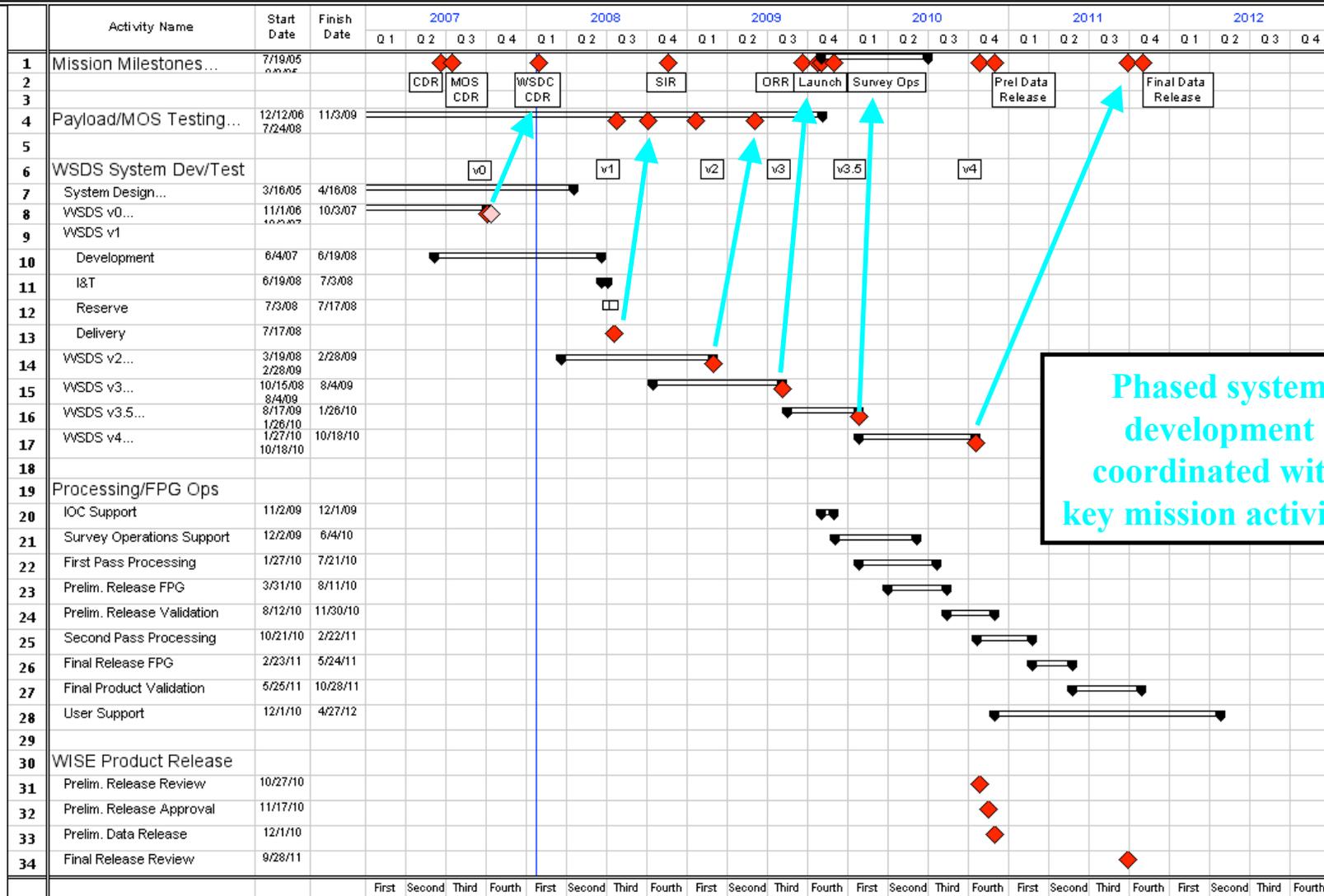




WSDC Top Level Schedule



WSDC Overview



Phased system development coordinated with key mission activities

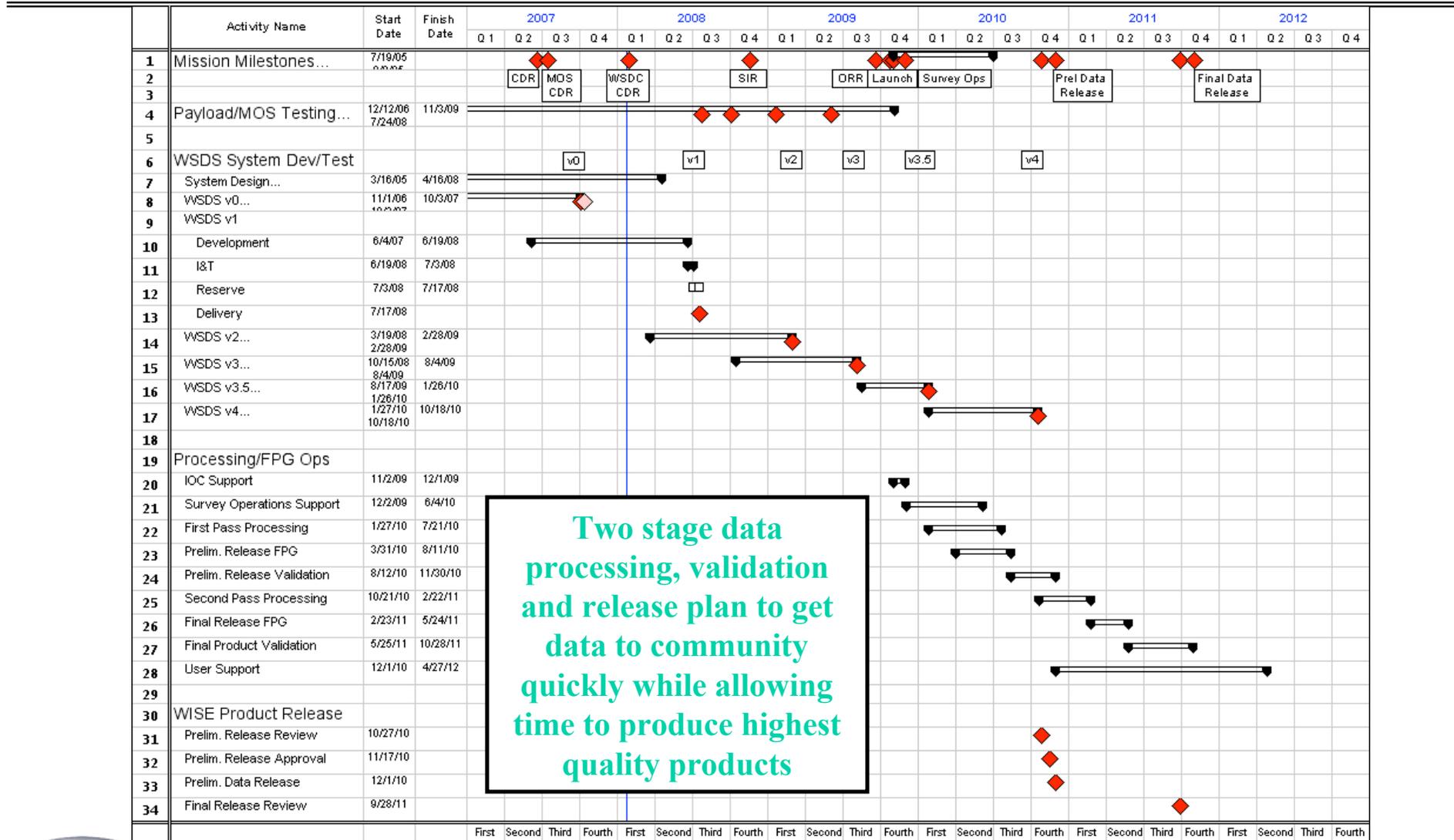




WSDC Top Level Schedule



WSDC Overview



Two stage data processing, validation and release plan to get data to community quickly while allowing time to produce highest quality products

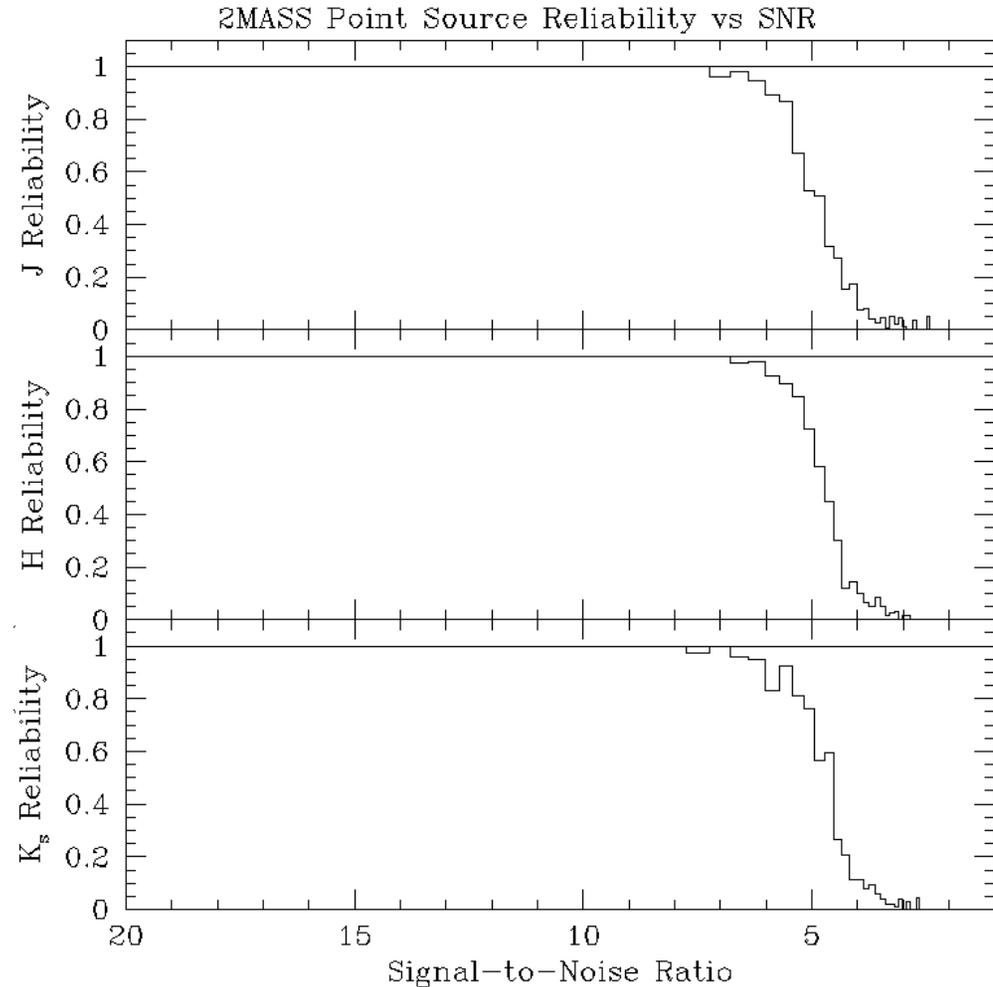




Follow 2MASS Approach for Achieving Catalog Completeness and Reliability



- WISE Source Catalog Reqs:
 - Completeness >95% for SNR>20
 - Reliability >99.9% for SNR>20
- Achieve completeness in MDET by detecting “sources” down to very low SNR levels (SNR~3)
 - Produce a source *Working Database (WDB)* that contains real sources, low SNR noise excursions, detections of artifacts, etc.
- Achieve reliability in FPG step by selecting subset of WDB detections using criteria from source characterization (WPHOT) and artifact flagging (ARTID)
 - $P(rel) = f(\text{SNR, bands-detected, profile-fit } \chi^2, \text{ confirmation stats (N/M), artifact flagging, etc.})$

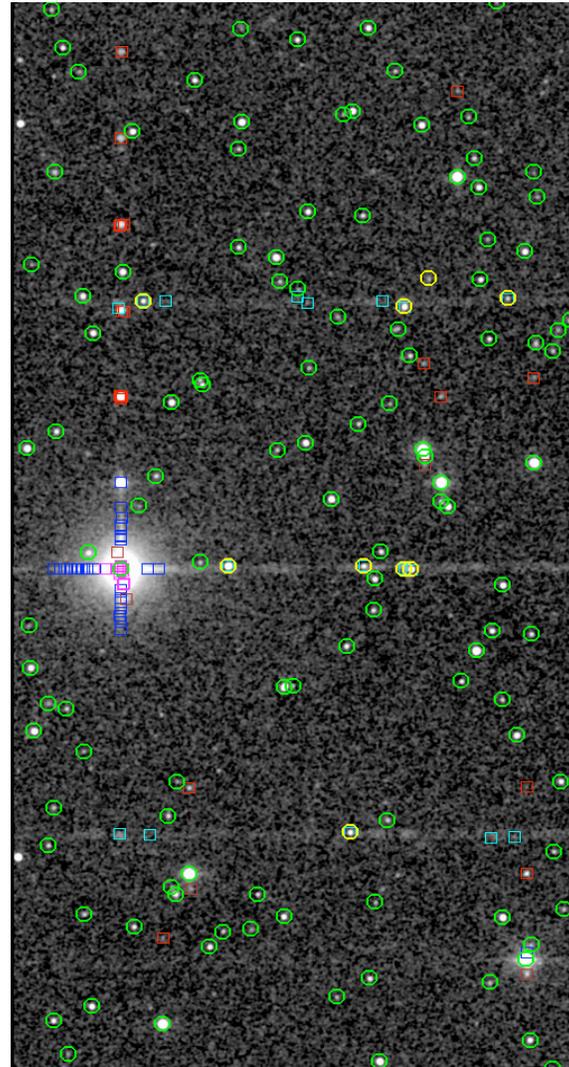




Artifact Identification



- Identify and flag in the extracted source *Working Database*:
 - Spurious detections of image artifacts such as latent images, diffractions spikes, optical and electronic ghosts produced by bright stars
 - Real sources that may be affected due to proximity to an image artifact
- Flagging based on proximity to predicted position/brightness of artifacts relative to bright “parent” sources
 - Based on successful 2MASS algorithms
- Design/implementation to begin March 2008 (WSDS v1) following start of payload ground characterization



Artifact Identification on
2MASS Atlas Image:

- Green** - True sources
- Yellow** - True sources contaminated by artifacts
- Red** - Latent images
- Blue** - Diffraction spikes
- Magenta** - Dichroic glints
- Cyan** - Electronic ghosts

Not shown: low SNR
noise extractions





Final Product Generation (FPG)



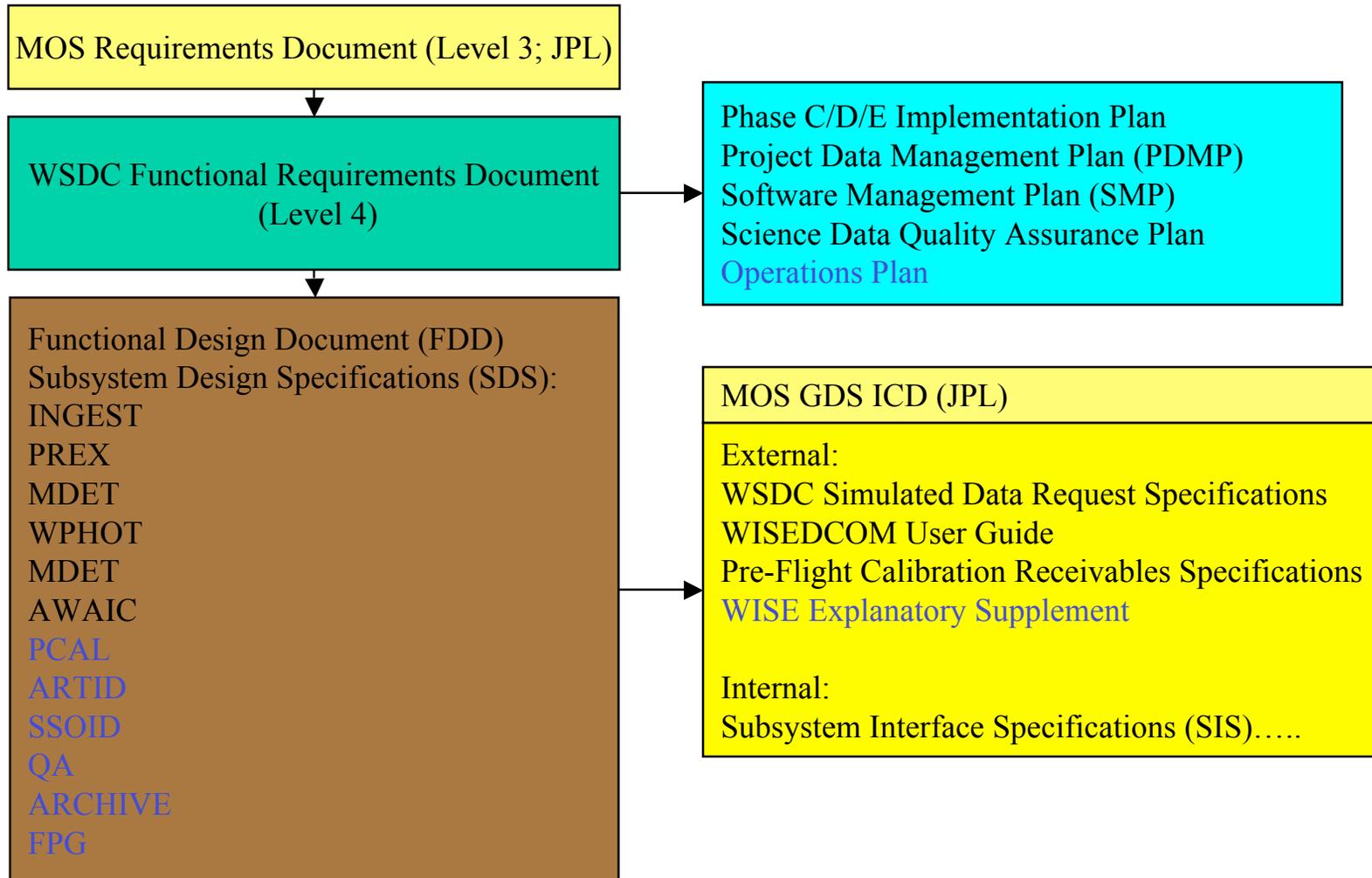
Human analysis and database-intensive activity to generate release Source Catalog and Image Atlas from L3 source Working Database and Coadded Image archive (not a discrete software subsystem per se)

- Source Catalog tasks:
 - Select subset of entries in WDB that satisfies reliability criteria
 - Define final Catalog schema and derive new columns, if necessary (Science Team discussing now)
 - Validate Catalog quality relative to L1/1.5 requirements
 - Document
- Image Atlas tasks:
 - Define final FITS headers, image metadata and derive new parameters, if necessary
 - Validate Image quality relative to L1/1.5 requirements
 - Document
- Tailored to characteristics of data and actual system performance
 - Extensive use of processing QA output and off-line analysis
- Design/implementation to start Jan 2009 (WSDS v3)





Document Tree





Testing Hierarchy



- System testing (*see Conrow/Testing presentation*)
 - Unit testing led by cog-eng/sci for each module
 - Delivery I&T, verification for major WSDS deliveries
 - Pre-launch system verification to the extent represented by test data
 - Test data
 - Simulated data from SOC/UCLA
 - Sky data from Spitzer archive and special WISE Team DDT observations
- Product Validation (*see Kirkpatrick/QA presentation*)
 - On-orbit processing output QA relative to science-based metrics
 - Characterization of output products relative to Mission Science Requirements
- IOC activities (*IOC Plan peer review 3/08 (TBS)*)
 - Support operational preparation (e.g. scan rate synchronization, annealing)
 - Update calibration products
 - Tune pipeline parameters to on-orbit performance





Archive Hierarchy



- Ground Test Data archive
 - Password-protected ftp access for Project/Science Team
 - Operational 12/07
- Operations archive (*see Conrow/Design presentation*)
 - Level 0-3 image and tabular data produced/access for WSDS, QA
 - Spinning disk integrated into processing system
- Level 0 data archive
 - Offsite tape copies for emergency recovery
 - Delivered to NSSDC following end of mission
- MOS engineering products archive
 - Backup storage of engineering telemetry, derived products for emergency recovery
- Data Products archive (*see Berriman-Groom/Archive presentation*)
 - Processed image/database access for Project Team, public
 - Integrated in NASA/IPAC Infrared Science Archive infrastructure





Action Item Status



Review	Item	Status	Response
MOS PDR	Simulated image data for testing software development is almost invisible in the plans.	CLOSED	Simulate data specifications and schedule document signed by SOC, WSDC and MOS
MOS PDR	Not clear when WSDC Ingest-1 s/w is capable of ingesting flight formatted packets and producing FITS images (unless its 9/2008 which is very late in the GDS development schedule)	CLOSED	WSDC HRP data prototype ingest system delivered 2/2007 to support MUB testing. Generates FITS images from HRP compressed packets.
MOS Peer CDR (finding)	WSDC delivery late with respect to MST	CLOSED	WSDC deliveries are properly synchronized with relevant MSTs
MOS CDR	Develop a plan to archive all raw images either on physical media at White Sands or by using IPAC capabilities as part of the data ingest processing.	CLOSED	WSDC will archive all Level 0 data as part of ingest processing
Mission CDR	WSDC should make contact with WFC3 development team to get information on their software development for the HgCdte detectors.	OPEN	Met with WFC3 IST lead (H. Bushouse). Agreed to hold TIM following WFC3 thermal vac.





WSDC Concerns



- Sound ground test plan in place, but responsibility for generation and delivery of pre-flight calibration products to WSDC is still being negotiated
Mitigation: Product generation and delivery plan in active discussion at project level.
- Surprises in actual on-orbit instrument performance
Mitigation: Modular S/W system with extensive parameter file control. Close involvement with payload ground test design discussion. Take advantage of lessons learned from Spitzer (Si:As). Lien for additional instrument characterization scientists.
- Handling deliveries of fragmented, out-of-order, and/or incomplete data frames
Mitigation: Design for expected fragmentation. Working with MOS to insure that sufficient information to reassemble fragmented files will be provided.
- Aggressive data release schedules allows very little time for validation
Mitigation: Relaxed requirements for Preliminary Release. Design for automated QA concurrent with processing, tied to science metrics linked to key science requirements (2MASS heritage).





Summary



- WISE data processing, archiving and distribution tasks are challenging, but we know how to do them
- The WSDC processing and archiving system design is based closely on proven systems
 - Scope and functionality of WISE data processing, quality assurance, final product generation and archiving tasks are similar to those of 2MASS.
 - 2MASS photometric/astrometric accuracy, completeness and reliability requirements were more stringent than those for WISE, and all were met or surpassed.
 - WSDC Level 0 data ingest system design draws upon operational GALEX system (not strictly an IPAC institutional activity, but supported by IPAC staff detailed to GALEX)
 - Leveraging existing infrastructure of the NASA/IPAC Infrared Science Archive
- Staff ramp-up, design and development proceeding on schedule and on-budget
- Key challenges are identified

