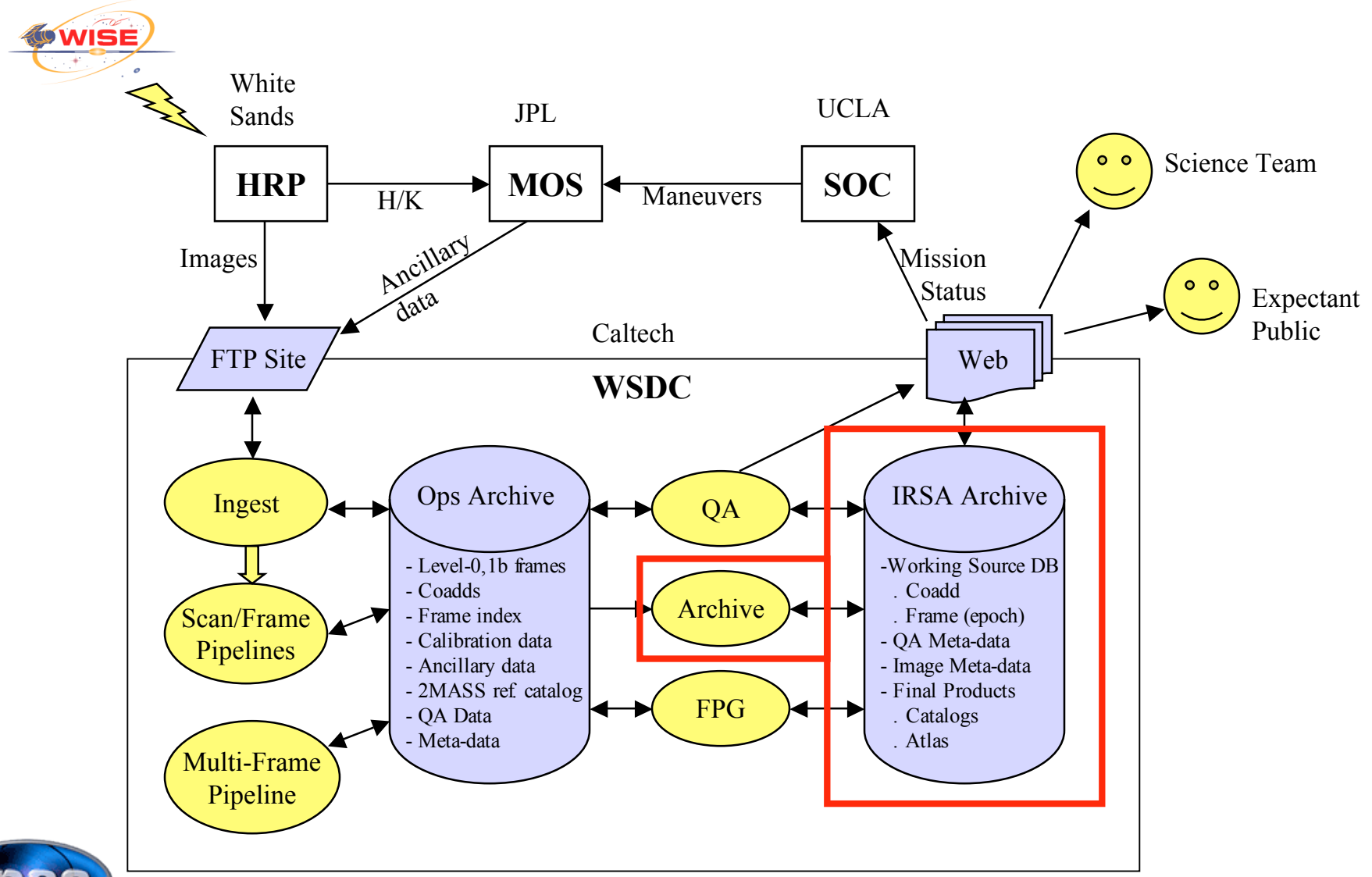
WISE Archive

Bruce Berriman / Steve Groom
Infrared Science Archive (IRSA), IPAC/Caltech
Roc Cutri
IPAC/Caltech

WSDC Functional Block Diagram



Driving Requirements - 1

Excerpted from WISE Science Data Center Functional Requirements, v2.0, 25 Nov 2007:

ID	Requirement	Traceability	Verification Method	Notes
	2.2.3 Archive			
L4WSDC-050	The WSDC shall create a copy of the Level 0 science data in a medium appropriate for permanent long-term storage.	L3MOS-278	Demonstration	
L4WSDC-051	The WSDC shall make the WISE catalog and image products available to the community via the internet through appropriate web-based tools.	L3MOS-383	Demonstration	
L4WSDC-052	As a goal, the WSDC will maintain the data products in a way that distribution of the complete WISE Source Catalog to users via portable media would be possible.	L3MOS-385	Inspection	
L4WSDC-053	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 data processing phase, and to insure interoperability with other NASA mission archives.	L3MOS-387	Inspection	
L4WSDC-054	The WSDC shall maintain a complete copy of the WISE science data set and software source code at a secure off-site location during the WISE mission to ensure survivability in case of major catastrophe.	L3MOS-389	Inspection	
L4WSDC-056	The WSDC shall maintain an archive of metadata derived from data processing for the individual science images for the duration of the project for the purpose of analysis and support of image access tools.	Self-derived	Demonstration	

Driving Requirements - 2

ID	Requirement	Traceability	Verification Method	Notes
	2.2.3.1 Data Access			
L4WSDC-060	The WSDC archive shall provide a web-based interface to enable selection, display and retrieval of any or all single-epoch images and combined Atlas Images based on position or time of observation for the purpose of quality assurance, validation and analysis. The goal shall be to also allow image selection on any image metadata parameter.	Self-derived	Demonstration	
L4WSDC-061	The WSDC archive shall provide a web-based interface to enable selection of sources extracted from single-epoch frames and/or combined Atlas Images based on position, flux, or combinations of any parameter maintained in the extracted source databases or Source Catalog.	Self-derived	Demonstration	
L4WSDC-086	The web-based interface to the WISE Image Atlas shall allow the user to view and retrieve an image in any of the four WISE bands with any specified center (tangent point) and any size up to at least 1ox1o.	Self-derived	Demonstration	

Summary of functionality

- Provide interfaces for regular (daily-weekly) automated ingestion of WISE products
- Archive of source databases, catalogs, and metadata tables
  - Test data, operations data, long term archival data
- Provide data access services for science team and general public
  - Catalog search
  - Image retrieval
  - Image mosaics, cutouts
  - Multi-color image comparison

IRSA overview

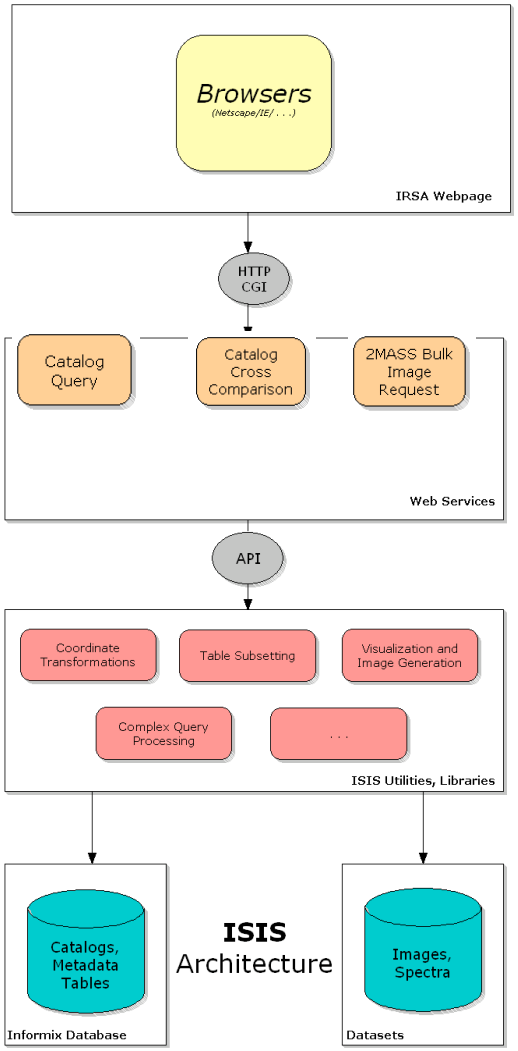
- IRSA is NASA's Archive node for IR and Submm data sets
  - Opened for business in 1999, and today has over 60 years combined archive experience on staff
  - Provided database management support and product generation for 2MASS
    - Driver for highly extensible and scaleable hardware and software architecture
  - Now serves data from 18 projects and missions & hosts seven contributed data sets
  - Physical archive hosts over 200 source catalogs, 11 million images, and 100,000 spectra
  - Archive services in wide use in the community
    - Web and program interfaces to deliver the full science content of the data sets, and maintains the *Montage* mosaic engine source code
    - Over 1000 peer reviewed papers cite IRSA or its services
    - Web page received over 12 million hits in 2007 (up x2 from 2007) and 32 TB of data downloaded
  - Over 60 years combined archive experience on staff
  - Interoperable with major archives
    - Member of National Virtual Observatory collaboration

IRSA's Hardware Architecture

- IRSA uses high-end servers and mass storage to curate the data products with built in fault-tolerance and redundancy, and provide performance and throughput
  - 4x redundancy in file storage
    - Full copies of data on *two* independent disk farms
    - Each disk farm is configured at RAID level 5
  - All hardware placed under maintenance contracts to ensure 24 x 7 operations
  - Database tables backed up regularly by IRSA on dedicated tape farm in ASCII and binary formats
  - Hardware replacement interval is 3-4 years
  - **IRSA has never suffered permanent data loss in its operational lifetime**

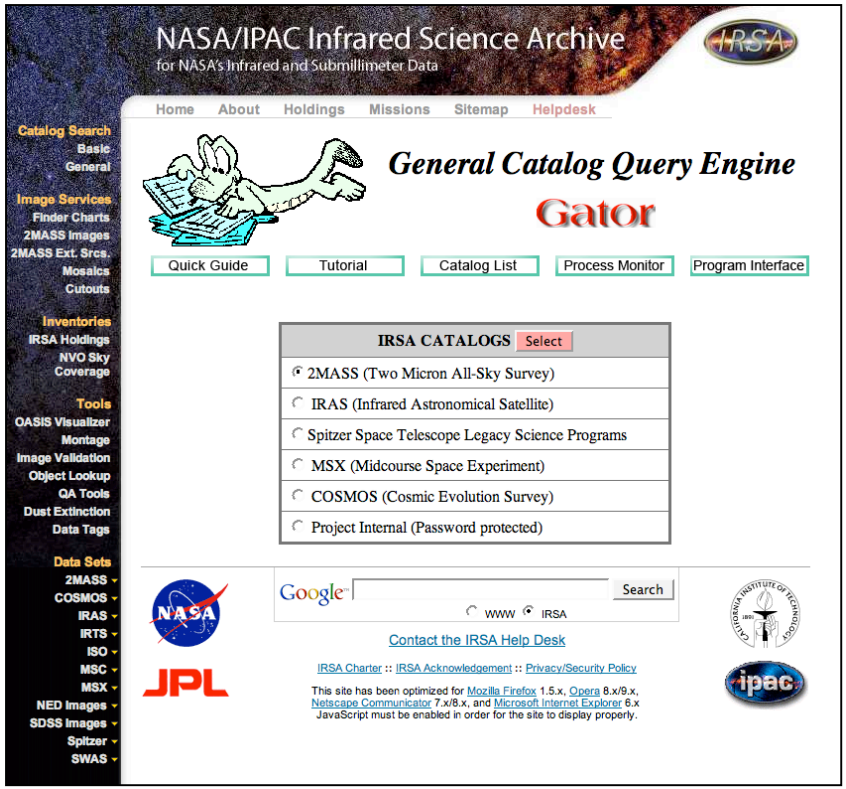
IRSA's Software Architecture

- WISE will inherit a mature, operational software architecture
  - Developed to provide archiving support to 2MASS
  - Re-usable, component based architecture allows efficient development of new services
  - New services are "thin" front ends sitting atop the Infrared Science Information System (ISIS)
  - Catalogs and metadata tables reside in Informix database



IRSA Catalog Services

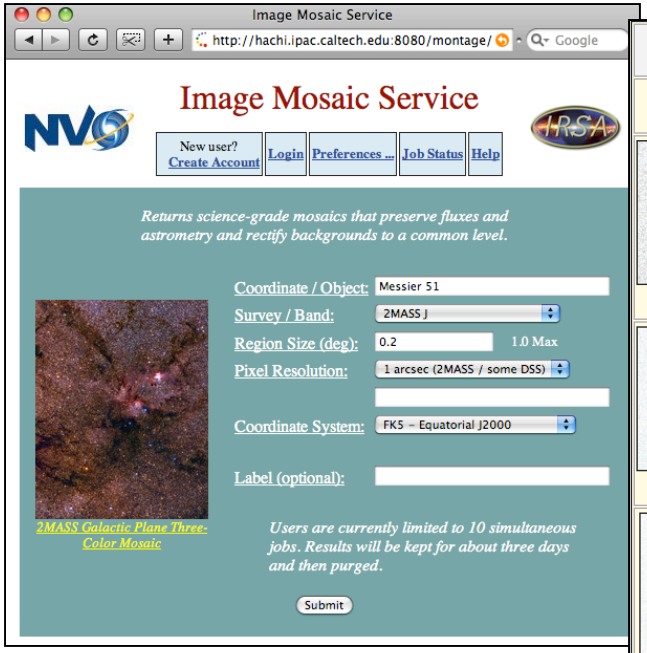
- General query tool (Gator)
  - Constraints against all catalog fields
  - Cross-matching of uploaded source tables against catalogs
  - Unique capabilities in astronomy
- Quick access to IRSA's most popular catalogs (BabyGator)
- *Catalog services will naturally include WISE catalogs*
  - User interface built dynamically

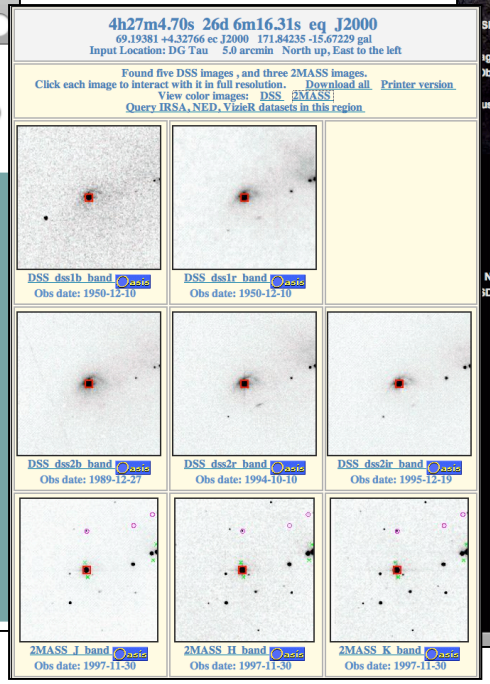


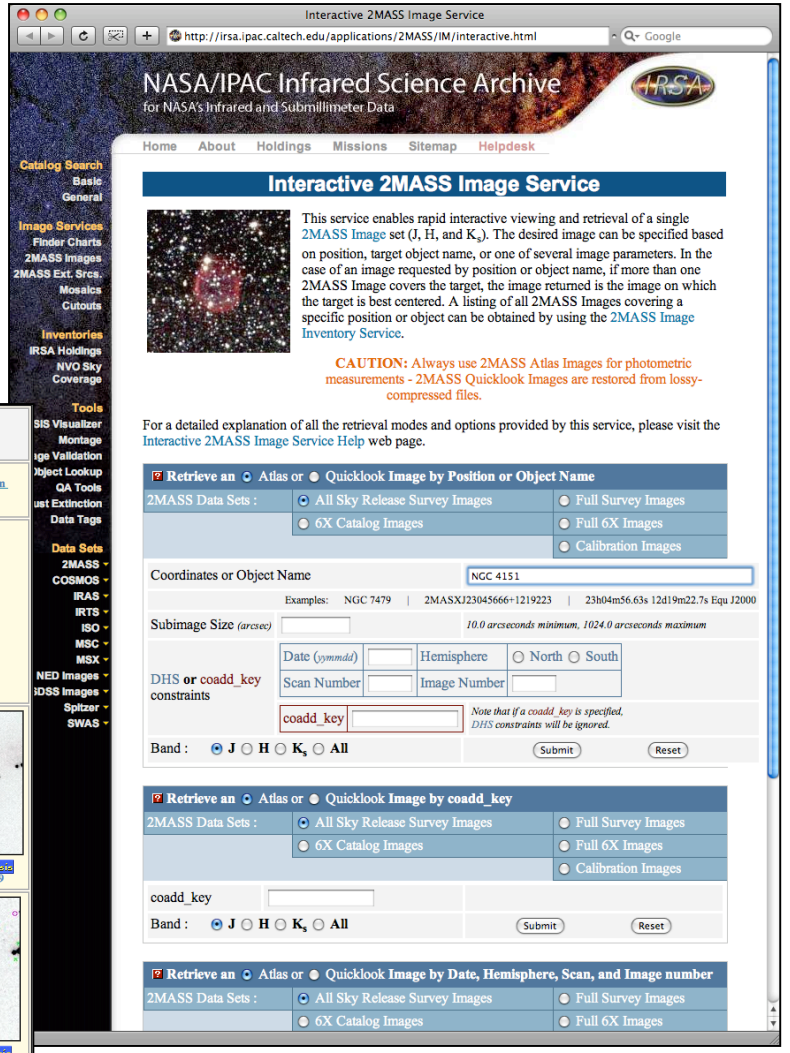
http://irsa.ipac.caltech.edu/applications/Gator/

IRSA Image Services

- Image Retrieval - interactive search of image holdings
- Image Mosaic - generate custom mosaics, subimage cutouts, and multicolor cross-comparison images
- Finder Chart - create catalog overlays and multicolor cross-comparison images from various missions and epochs
- *WISE Atlas image products (including co-adds) will be available through these services*

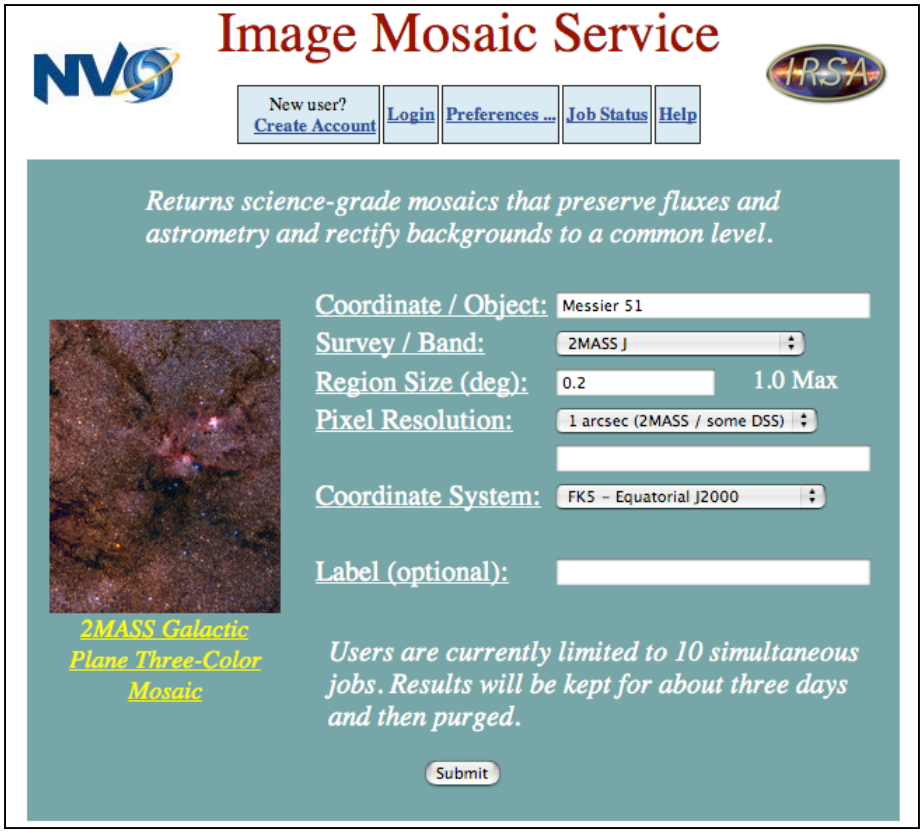






Co-add Pixel Query Service

- Atlas Image products will be accessed via a WISE image server interfacing with IRSA. Users can retrieve a co-add *portion* based on sky location, spatial extent, orientation and pixel scale.
- Maximum query size will be at least $1° \times 1°$ (constrained by server load and algorithm - TBD).
- Note: IRSA already provides a mosaicking and an image cutout service, e.g:



IRSA Services for WISE

- Periodic delivery of WISE products into IRSA (~20GB/day)
  - Source DB's, metadata, images
  - Software Interface Spec (SIS) for transfers to be developed
- WISE catalogs will be searchable through Gator
- Image services will provide access to original WISE L1B and L3 products
- Mosaic and cutout services will provide custom images with full traceability to original WISE products

Development Schedule

- Roc ?

Issues/Concerns

- IRSA has not handled throughput as high as that from WISE
- Benchmark tests indicate that the current architecture will scale to the throughput of WISE
  - 2MASS All Sky Catalog used as tesw catalog
  - Load times into Informix
    - 5 mins to load 10 million rows
    - 20 mins to load 50 million rows

    ⇒Will be able to handle the load from WISE
  - But indexing times depend critically on table size
    - E.g. to create four indices:
      - 10 millions row takes ~8 minutes
      - 50 millions row takes ~65 minutes
      - 1.3 billions rows takes 12 hours