IRSA Support for WISE Science Data Center Operations

IPAC Director's Review
20 October 2009

Harry Teplitz
Steve Groom

Requirements

Requirements taken from the *WSDC Archive Design Document (v1.3)*

- Operations: IRSA will provide image and catalog services for use by the WSDC and the WISE Science Team
  - *Ingest large data deliveries (images, catalogs, meta data) twice per week, and serve to Science Team members*
- Post-flight: At the completion of the WISE project activities, IRSA will assume curation of the data products and documentation
- WISE data will be made accessible to cross-mission discovery services within IRSA, and to standard program interfaces needed for inter-operability with other NASA archives
- NEO-WISE
  - *Archive Level 1b images and sources*
  - *Develop an enhanced interface to IRSA services that would support solar-system specific queries using the single-epoch WISE image and source data.*

Required WISE Services at IRSA

- Catalog search interface
  - *Allow searches of Level 1b and Level 3 sources, using constraints on data values (position, flux, flux ratio)*
  - *Enable queries on any meta-data table in the Archive*
  - *Enable complex queries*
- Image Service
  - *Search for, display, and retrieve Single Epoch and Atlas images*
  - *Search by position, time*
  - *Enable batch requests*
- Image Pixel Server
  - *Retrieve Atlas images in any band with size up to 2x2 deg.*
  - *Retrieve comparison data from other sky surveys (Finder Chart Mode)*
  - *Return both FITS and JPG*
- Image Inventory Service
  - *Provide listing of all WISE images (and metadata) in specified region*
- NEO-WISE
  - *Efficient search algorithm and user-friendly interface*

Technical Challenges

- Very large tables drive database complexity
  - *WISE L1b sources estimated at 20-30 billion rows*
  - *Comparison: 2MASS PSC = ~0.5 billion, USNO-B ~1 billion*
- Frequent (twice weekly) deliveries require rapid table ingestion and indexing
  - *Not enough time between deliveries to complete a full re-indexing*
  - *Need to sustain performance while adding new data*
- Incremental ingestion necessitates rework of spatial indexing methods and how search services use them
  - *Need to update spatial indexing incrementally*
- Moving object searches require new algorithms
  - *Need spatial+temporal indexing to match position at time*
  - *Need to incorporate mechanism for orbit propagation into search functionality*
- Image services require custom interface
  - *Project requirements are unique enough that they are not easily folded into existing services*

Methodology

- Data ingest
  - *Load separate table for each incremental delivery of L1b sources*
  - *Multi-table "view" avoids re-indexing of entire catalog on each delivery*
  - *New spatial indexing algorithm supports incremental deliveries*
- Catalog service
  - *Re-use existing IRSA catalog DB infrastructure and interface ("Gator")*
  - *Password protection for science team access to WISE products*
- Image Services
  - *Custom interface for science team access*
  - *Image services will be folded into IRSA upgraded interface by end of 2010 for use in public release*
    - Maintains functionality of custom interface, but at lower maintenance costs in the future
- Moving object searches
  - *New, efficient search algorithm*
  - *Re-use of IPAC or JPL software for orbit propagation*
  - *Need to develop new spatial+temporal image indexing*
  - *Interface to be part of IRSA upgrade*

Status

- Data ingest
  - *Multi-table "view" demonstrated; enables ingest w/o degrading performance*
  - *Prototype indexing to be completed week of 10/19, integration with Gator pending*
  - *Development of data transfer scripts still pending*
- Catalog service has been demonstrated with simulated data
  - *Largest catalogs served by IRSA to date, with adequate performance on development hardware*
  - *Data update expected in week of 10/19*
  - *Data update expected in November with TBD schema changes*
- Image service
  - *Prototype 1 – May, 2009*
  - *Prototype 2 – July, 2009*
  - *V1 – delivered to WSDC on 10/9*
  - *V1.1 – due Dec 1, 2009*
    - Need to prioritize requested changes against other IRSA activities (especially support of spatial indexing of incremental deliveries)
  - *V2 – anticipated end of 2010*
    - Change over to upgraded IRSA image search interface
- Moving Objects
  - ~~*New hire (Parades) working w/ WSDC team on software and algorithm*~~

Systems

- File server & storage *
  - *New IRSA file server dedicated for WISE*
  - *IRSA standard: Sun X4200 server, Nexsan disk arrays*
- DB server & storage *
  - *Will keep WISE on Informix through FY11, migrating along with the rest of IRSA after that time*
  - *WISE databases to be hosted on otherwise-idle IRSA "standby" server.*
    - No provision for failover of WISE DB services – data size makes that impractical
  - *IRSA standard: Informix running on Sun T5240*
- Server sizing
  - *DB*
    - Hardware in place for L1b=25B rows, 9mo mission (= ~45TB)
  - *Image services*
    - Hardware in place for ~96TB image data, expansion anticipated late FY10

* Note: WISE is paying for IRSA hardware expansions as needed to support WISE

Systems - 2

- Web Server
  - *WISE web apps to be hosted on dedicated web server*
    - Gator, image service – separate from main IRSA services
- Data transfer server
  - *Manages transfer of data from WISE ops to IRSA, working within negotiated time and data rate windows*
  - *Copying of WISE products to IRSA servers avoids direct exposure of performance-sensitive WISE ops systems to unpredictable user loads*
- Backups – IRSA/ISG standard
  - *Utilizing IPAC shared backups system, including offsite storage*
  - *Will need to organize data layout to avoid unnecessary duplication of backups*

Performance Testing

- Testing of incremental catalog loading techniques indicate that the twice-weekly loads should be doable in ~24hours.
- Efforts underway to confirm performance using operational configuration
- Have identified some areas for improvement in loading speed, to be pursued as necessary

Dependencies

- IRSA needs the following from WSDC:
  - *Finalized schema for flight operations phase, especially for L1b sources*
    - Because of L1b data size and ingestion rate, it will be nearly impossible to make changes to catalog schema once regular deliveries commence
    - Expected ~11/4
  - *Specs and implementation for ops data transfer*
    - Detailed list of data products, meta-data, and formats
    - Negotiate mechanisms to trigger and manage transfers
    - Expected by ~11/4
  - *Agreement on prioritization of image service upgrades*

Mission Support

- IRSA will support WISE activities in parallel to core functions, Planck support, and Spitzer closeout activities
  - *Staffing separation*
    - WISE support: Monkewitz, Zhang, Terek (w/ Rey and Groom)
    - SHA activities: Wu, Roby, Loi, Balandran, Teplitz, Howell (w/ Groom)
    - Planck support: Mi, (w/ Teplitz, Groom)
    - Core activities: Mi, Terek, Rey, Groom (w/ Teplitz, Howell)
  - *Hardware separation*
    - WISE and SHA are supported by independent servers and storage to minimize interference
    - IRSA will serve WISE products with minimal impact to WISE ops servers
  - *Database separation*
    - IRSA uses separate DB servers for IRSA+NExScI, WISE, SHA
    - Rey is shared between IRSA, WISE and NExScI
    - SHA DB functions supported by SSC SDM team
- IOC support
  - *No direct support for WISE IOC activities expected from IRSA*
- Public data release and post-flight curation are part of design requirements for IRSA infrastructure upgrades