



The IPAC Research Archives

Steve Groom
IPAC / Caltech





IPAC overview



The Infrared Processing and Analysis Center (IPAC) at Caltech is dedicated to science **operations**, data **archives**, and community **support** for astronomy and solar system science missions, with a historical emphasis on infrared-submillimeter astronomy and exoplanet science.

IPAC is entering its 30th year of building and operating successful data centers and research archives for space- and ground-based astronomical observatories and large-scale survey programs.

IPAC Activities

- NASA/IPAC Infrared Science Archive (IRSA)
- NASA Extragalactic Database (NED) – “Google for galaxies”
- NASA Exoplanet Science Institute (NExSci)
 - Exoplanet Archive
- Keck Observatory Archive (KOA)

IPAC is also home to:

- Spitzer Science Center
- NASA Herschel Science Center
- NASA/US Planck Data Center
- Las Cumbres Observatory Global Telescope (LCOGT) Archive



IRSA Overview

IRSA is NASA's IR/sub-mm archive

- Began with IRAS and 2MASS, leading into the decade of IR missions
- Spitzer
- Wide-Field Infrared Survey Explorer (WISE)
- NASA Planck Archive

The screenshot shows the NASA/IPAC Infrared Science Archive website. At the top, it says "NASA/IPAC Infrared Science Archive for NASA's Infrared and Submillimeter Data". There is a navigation menu with "Home", "About", "Holdings", "Missions", "Documentation", and "Help Desk". A search bar is prominently displayed with the text "Search IRSA with the Data Discovery Service". Below the search bar, there are examples of object names and coordinates, and a search radius input field. On the right side, there are several "New" items listed, including "Finder Chart v2 beta", "WISE Image Service", "NASA Planck Archive", and "Spitzer Heritage Archive". The "News and Updates" section features a date of "April 18, 2013" and a headline about the release of new data from the S4G Spitzer Exploration Science program and the COSMOS survey. Below this, there is a sub-headline "The S4G catalog is now available, including photometry, model fit parameters, measurements from the literature, and links to the S4G data for each galaxy. Ellipse fit models are also now available for all 2,352 galaxies." Another update dated "March 21, 2013" mentions the first release of all sky temperature maps, foreground component maps, and cosmology results from the Planck mission.

IRSA also provides access to many other mission datasets, including some hosted at other institutions, brought together for the convenience of researchers



Purposes behind archiving

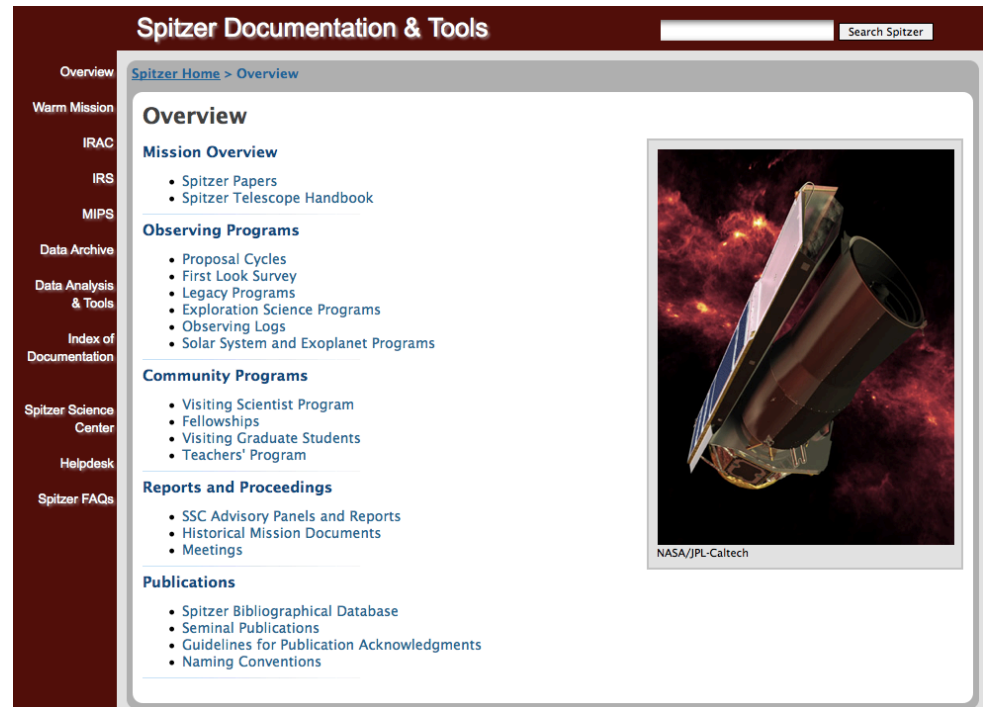
- Permits others to exploit data – now and into the future
- Provides efficient access to mission data to *enable research that has not yet been envisioned*

The most important questions our users want answered are:

- “*How do I get (and use) my data?*”
- “*Whom do I ask if I have a problem?*”

Archiving involves:

- Data curation
- Data access
- Documentation
- Tools for data reduction, analysis
- User support



The screenshot shows the 'Spitzer Documentation & Tools' website. The page has a dark red header with the title 'Spitzer Documentation & Tools' and a search bar. A left sidebar contains a navigation menu with links to Overview, Warm Mission, IRAC, IRS, MIPS, Data Archive, Data Analysis & Tools, Index of Documentation, Spitzer Science Center, Helpdesk, and Spitzer FAQs. The main content area is titled 'Overview' and includes sections for Mission Overview, Observing Programs, Community Programs, Reports and Proceedings, and Publications. A large image of the Spitzer Space Telescope is on the right side of the page.

Spitzer Documentation & Tools

[Spitzer Home](#) > Overview

Overview

Mission Overview

- Spitzer Papers
- Spitzer Telescope Handbook

Observing Programs

- Proposal Cycles
- First Look Survey
- Legacy Programs
- Exploration Science Programs
- Observing Logs
- Solar System and Exoplanet Programs

Community Programs

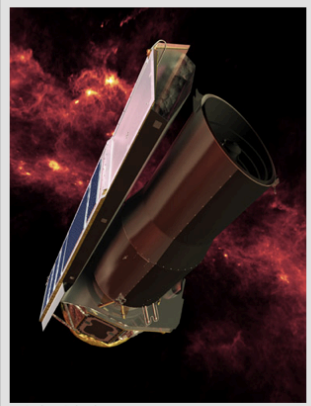
- Visiting Scientist Program
- Fellowships
- Visiting Graduate Students
- Teachers' Program

Reports and Proceedings

- SSC Advisory Panels and Reports
- Historical Mission Documents
- Meetings

Publications

- Spitzer Bibliographical Database
- Seminal Publications
- Guidelines for Publication Acknowledgments
- Naming Conventions



NASA/JPL-Caltech

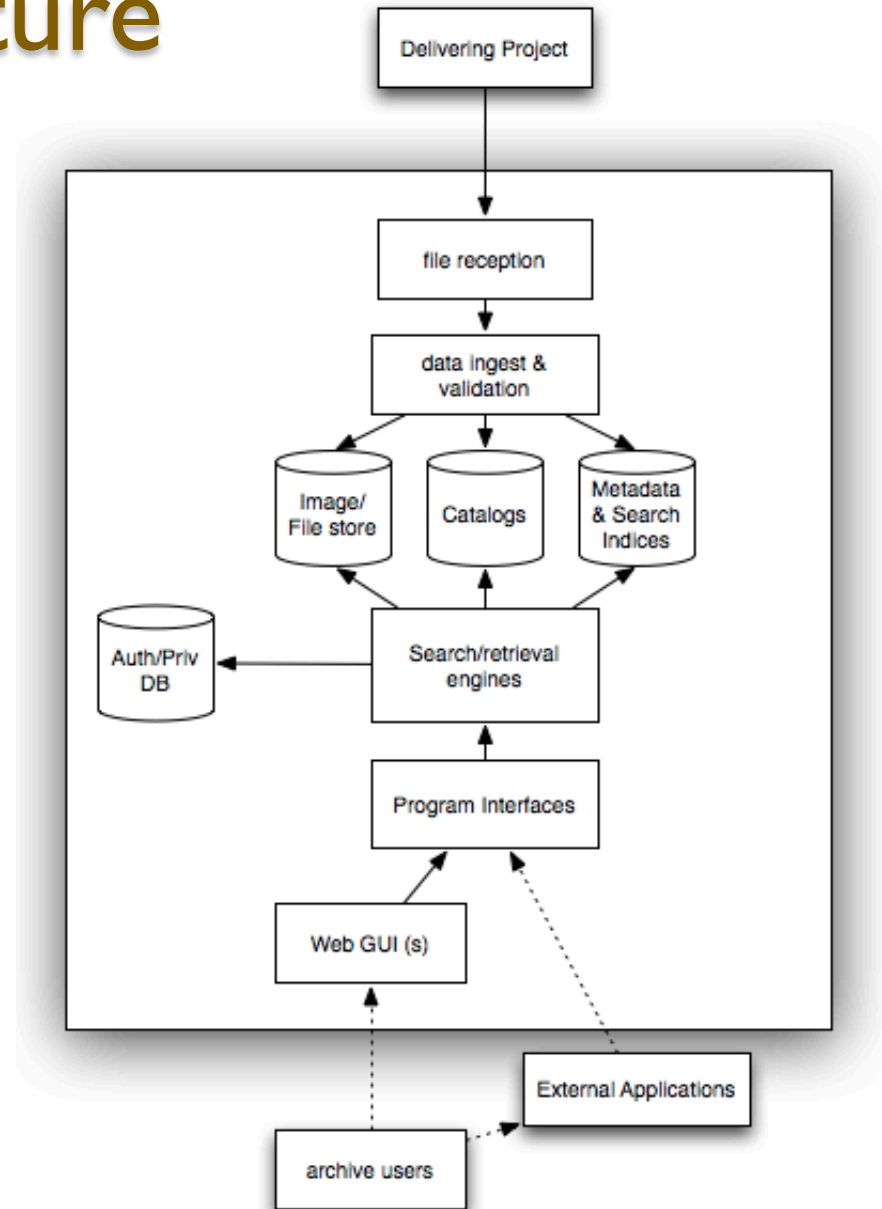
All are important to ensure usefulness of the data

Data access models

- **Find:** Discovery/Inventory– what’s available?
- **Query:** Detailed, dataset-specific query
 - Find records/images matching constraints
- **Combine:** Multi-dataset query
 - Display data from multiple datasets together
- **Mine:** Bulk query
 - Upload list of search positions or other criteria
- **Retrieve:** Bulk dataset download
 - Give me the whole thing
- **Remote Access:** Direct external access
 - “Can I just have my program connect directly to your database?”

Archive Architecture

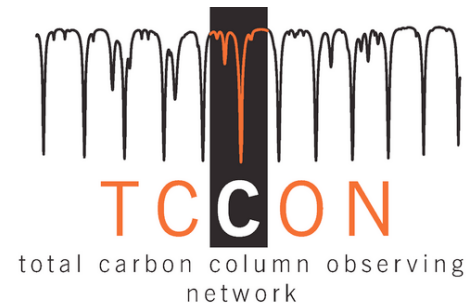
- Ingestion & Validation
- Storage, Indexing
- Reusable search engines
 - Catalogs
 - Image/extended spatial
- Product retrieval
 - Also on-demand generation (e.g. image cutout, mosaic)
- Program (VO) interfaces
- Web UI's
- Support for external UI's





IRSA's common architecture supporting other activities

- NExSci – Exoplanet Archive, Kepler Science Analysis System
- Keck Observatory Archive
 - 20 years of observations
- Solar System – NEOWISE
- Non-NASA:
 - P60, PTF
 - LCOGT
 - LSST
 - TCCON (Atmospheric research)



Questions to ask in archive design

- What are the products?
 - Standard levels of processing/calibration/reduction?
- What are the use cases for search&retrieval?
- What are the units for data packaging, query, retrieval?
- What can be queried?
 - Metadata – date, instrument, position, etc
 - Pre-calculated summary metrics/statistics
 - Measurement data
 - Derived/calculated properties, combinations of data
- Any proprietary data considerations?
 - Mixing public and private data within the archive can greatly complicate handling of queries involving summary information



Questions - 2

- Who needs access to the data?
- When do they need access?
 - Real or Near-real time – notification, event distribution: a topic all to itself!
 - Hours/Days/Months/Years later
- Acceptable query latency?
 - Interactive vs batch-style queries
- Anticipated usage/volume

Things to think about

- Consistency of formats & organization
- Completeness/Correctness of data
 - initial validation is important
- Completeness of documentation
- Examples are helpful!
 - Data reduction “cookbooks”
 - User tutorials, workshops
- Important for the long haul:
 - Consider longevity of technologies, data formats, programming languages/systems
 - Media lifetime, periodic refresh



Trends: Rapid evolution of information-handling systems

- Technology for data handling changes even more quickly than technologies for photon gathering
 - Fast internet everywhere
 - Extremely portable disks and drives
 - Grid computing and protocols for remote analysis
 - The cloud

Trends: Archives as analysis environments

- Browse-and-download
 - Identify data of interest, take it home for further study
- Complex queries
 - Finding “interesting” data within large datasets
- In-database analysis?
 - More complex queries over larger data volumes
 - Data size growing faster than communications
 - Bringing the software to the data?
 - “ ‘Big Data’ means you can’t move it.”

Common Formats and Tools

- Catalogs – tabular data
- Images – usually as FITS files
- Spectra
- Time-series photometry
- light curves

Home About Holdings Missions Documentation Helpdesk

General Catalog Query Engine

powered by Gator

Quick Guide Tutorial Catalog List Process Monitor Program Interface

CATALOG SELECTION: 2MASS

| 2MASS All-Sky Release Database Select | | | | |
|--|---|-----------|-----------|-------------------|
| Selection | Descriptions | # Columns | # Rows | Information |
| <input checked="" type="radio"/> | 2MASS All-Sky Point Source Catalog (PSC) | 127 | 470992970 | i |
| <input type="radio"/> | 2MASS All-Sky Extended Source Catalog (XSC) | 423 | 1647599 | i |
| <input type="radio"/> | The 2MASS Large Galaxy Atlas | 88 | 655 | i |
| <input type="radio"/> | 2MASS All-Sky Survey Scan Info Read Me! | 68 | 59731 | i |
| <input type="radio"/> | 2MASS All-Sky Survey Atlas Image Info | 134 | 1373813 | i |

| 2MASS Survey Scan Working Databases and Metadata Select | | | | |
|--|---|-----------|-----------|-------------------|
| Selection | Descriptions | # Columns | # Rows | Information |
| <input type="radio"/> | 2MASS Survey Point Source Reject Table | 72 | 843988897 | i |
| <input type="radio"/> | 2MASS Survey Merged Point Source Information Table | 56 | 165942357 | i |
| <input type="radio"/> | 2MASS Survey Merged Point Source Link Table | 3 | 396697288 | i |
| <input type="radio"/> | 2MASS Survey Extended Source Reject Table | 387 | 943441 | i |
| <input type="radio"/> | 2MASS Survey Merged Extended Source Information Table | 79 | 406636 | i |
| <input type="radio"/> | 2MASS Survey Merged Extended Source Link Table | 3 | 960841 | i |
| <input type="radio"/> | 2MASS Survey Scan Info | 405 | 70712 | i |
| <input type="radio"/> | 2MASS Survey Atlas Image Info | 134 | 1626376 | i |

| 2MASS 6X Scan Working Databases and Metadata Select | | | | |
|--|--|-----------|----------|-------------------|
| Selection | Descriptions | # Columns | # Rows | Information |
| <input type="radio"/> | 2MASS 6X w/LMC/SMC Point Source Working Database /Catalog Read Me! | 71 | 24023702 | i |
| <input type="radio"/> | 2MASS 6X w/LMC/SMC Merged Point Source Information Table | 56 | 4771737 | i |
| <input type="radio"/> | 2MASS 6X w/LMC/SMC Merged Point Source Link Table | 3 | 12267173 | i |
| <input type="radio"/> | 2MASS 6X w/LMC/SMC Extended Source Working Database / Catalog Read Me! | 387 | 247091 | i |



Configurable user interfaces

The top screenshot displays the IRSA search results interface. It features a search bar, navigation tabs (IRSA, Mission, Archive Search, Related Data Archives, Tools & Services, Help), and a search history section. The main content area includes a 'Search Again' button and a 'Results' table. The table lists search results with columns for Band, RA (deg), Dec (deg), ra1, dec1, ra2, dec2, ra3, and dec3. A 'Download Options' section is visible on the left, and a 'Preview: Wise - 3 color' image is shown on the right.

| Band | RA (deg) | Dec (deg) | ra1 | dec1 | ra2 | dec2 | ra3 | dec3 |
|------|----------|-----------|---------|---------|---------|---------|---------|------|
| 4 | 10.5004 | 40.7530 | 11.1999 | 40.5883 | 10.2918 | 40.2244 | 9.8041 | |
| 1 | 10.5008 | 40.7533 | 11.1952 | 40.5927 | 10.2899 | 40.2293 | 9.8072 | |
| 3 | 10.5002 | 40.7544 | 11.1937 | 40.5873 | 10.2837 | 40.2298 | 9.8082 | |
| 3 | 10.6028 | 40.8459 | 9.9050 | 41.0095 | 10.8237 | 41.3700 | 11.2920 | |
| 2 | 10.6025 | 40.8466 | 9.9023 | 41.0050 | 10.8237 | 41.3705 | 11.2951 | |
| 1 | 10.6022 | 40.8470 | 9.9036 | 41.0040 | 10.8237 | 41.3705 | 11.2951 | |
| 4 | 10.6026 | 40.8473 | 9.8989 | 41.0083 | 10.8237 | 41.3705 | 11.2951 | |
| 3 | 10.6690 | 41.1426 | 9.9678 | 41.3053 | 10.8237 | 41.3705 | 11.2951 | |
| 2 | 10.6688 | 41.1433 | 9.9650 | 41.3008 | 10.8237 | 41.3705 | 11.2951 | |
| 1 | 10.6684 | 41.1437 | 9.9664 | 41.2998 | 10.8237 | 41.3705 | 11.2951 | |
| 4 | 10.6688 | 41.1440 | 0.0616 | 41.3041 | 10.8237 | 41.3705 | 11.2951 | |

The bottom screenshot shows a detailed view of a search result for 'PLCKER0330 G017.00+00.84'. It includes a 'Search Results' table with columns for name, flux, flux_err, iras_100bg, cmbsubtract, extended, and dates. Below the table, there are several Planck and WMAP frequency maps displayed in a grid. The Planck maps are labeled with frequencies: 30 GHz, 44 GHz, 70 GHz, 100 GHz, 143 GHz, 217 GHz, 353 GHz, and 545 GHz. The WMAP maps are labeled with frequencies: 22.8 GHz, 33.0 GHz, 40.7 GHz, 60.8 GHz, and 93.5 GHz. The interface also includes a 'Details' section with 'Notes' and 'Additional Information' fields.

Decreases the costs of setting up access to new data sets and the costs of long-term maintenance. Provides unified user experience.

Summary

- Preservation of science data products as well as lower-level products enhances future research
 - Can't predict all potential use cases
- Pay attention to formats, provenance, documentation
 - Future usefulness of the data depends on it
 - “Quality” can be subjective. One man's trash is another man's treasure.
 - Caveat Emptor
- Reusable building blocks can effectively support a wide range of use cases
- Distinct perspectives for rapid vs long-term utilization
 - Fast observation followup vs future data mining
 - Discovery vs research use cases