Enhancing Science Return During the Rapid Expansion of IRSA

H. I. Teplitz, S. Groom, W. Roby, and the IRSA Team

Infrared Processing and Analysis Center, Caltech, Pasadena, CA 91105, USA

Abstract.   The NASA/IPAC Infrared Science Archive is undergoing a rapid expansion in the portfolio of missions for which we serve data, including the recent additions of the Spitzer Heritage Archive, the WISE archive, and NASA's Planck archive. This expansion has led to more than an order of magnitude increase in data holdings within a few years, and the inclusion of catalogs up to 10 billion rows. As a result, IRSA has a special opportunity to grow its services to support not only individual IR missions, but to optimize the synergy between them. We discuss the ways in which we enhance the science return from multiple IR data sets, including data discovery, moving object "precovery", and image cross-comparison.

1.   The NASA/IPAC Infrared Science Archive (IRSA)

The NASA/IPAC Infrared Science Archive (IRSA; http://irsa.ipac.caltech.edu) is the infrared component of the NASA archives. IRSA is chartered to (1) curate and serve scientific data products from NASAs infrared and submillimeter projects and missions, (2) enable optimal scientific exploration of these data sets by astronomers, and (3) support planning for, operation of, and data set generation from NASA missions. IRSA offers efficient access to IR mission holdings, through image and catalog search tools.

IRSA is currently undergoing a major expansion - the largest in our history - from holdings dominated by the Infrared Astronomical Satellite (IRAS Neugebauer et al. 1984) and the Two Micron All Sky Survey (2MASS Skrutskie et al. 2006) to a much broader portfolio. Recent additions include:

- The Heritage Archive of the Spitzer Space Telescope (SHA Werner et al. 2004): the first time a Great Observatory archive has transitioned to long-term curation;

- The Wide-Field Infrared Survey Explorer (WISE Wright et al. 2010): NASAs mid-IR all-sky survey, destined to join IRAS and 2MASS as a benchmark of IR astronomy;

- NASA's Planck Archive, with an emphasis on foreground research – Planck Planck Collaboration et al. (2011) has spatial resolution, sensitivity and coverage comparable to IRAS in an otherwise unexplored spectral decade (300 μm 2 mm);

- Seamless access to Herschel data through IRSA data discovery services, utilizing the Virtual Observatory (VO) protocols
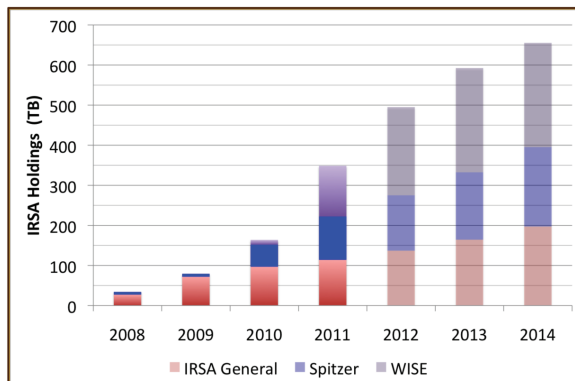
Figure 1.: The expansion of IRSA holdings by year. Shaded bars for 2012 onward are projected volume based on current projection.

IRSA is also undertaking a major revision of tools and services, especially the web portal and the database management system. By consolidating the code base for the many holdings, we will increase functionality and lower the cost of the services. This rapid expansion gives us the opportunity to greatly increase the science return of our services and datasets.

2. The Rapid Expansion

IRSA's expansion has resulted in a tremendous growth in the volume of astronomical data holdings. More than an order of magnitude expansion has already occurred, and the pace is expected to continue (see Figure 1). Together with the expansion in holdings, IRSA's suite of data analysis tools is expanding as well (Teplitz et al. 2012).

In addition to the volume of holdings, the complexity is also increasing. Catalogs are becoming much larger, and require significant new effort to maintain database performance. WISE, in particular has pushed the boundaries of IRSA's database management system (DBMS). The single-exposure database of sources released in the WISE All-Sky Data Release included about 10 billion rows. This table is the largest served to date by IRSA. It will be superseded by an even larger table associated with the ALL-WISE project (PI=E. Wright) in 2013. Furthermore, ingest of WISE and other catalogs requires systems to be designed to allow periodic ingest without interruption of service, as new data come in from operating missions.

Finally, there is a significant increase in the demand for science user support as IRSA holdings grow. Archives require active support by expert scientists (who use the data themselves) to enable new research by the community. As time passes, the complexity of user support questions grows, as the easy ones have already been answered.

3. Unifying Core IRSA Image Services

IRSA is improving its software architecture. We are developing a more robust system for efficient searches of images and catalogs. We are continuing to follow a software philosophy of keeping clean separation of the search layer and the presentation layer by
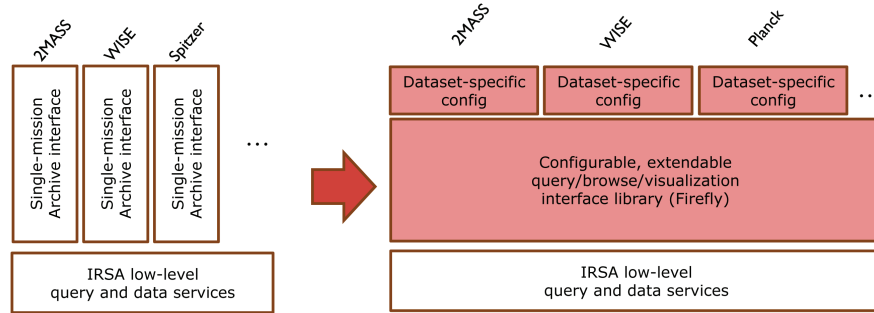
Figure 2.: Schematic of IRSA's software infrastructure. We keep a clean separation of the search and presentation layers, and we utilize a configurable interface ("Firefly"; see Roby et al. (2013)).

implementing low-level query services (see Figure 2). These services naturally lead to the program interface and our ability to implement VO-compatible searches and results.

Each mission requires some amount of custom configuration based on dataset-specific information. This can be accomplished either by writing a new interface for each mission, giving maximum customization, or by writing a flexible interface that will meet most of the needs. The latter approach has significant advantages for keeping the cost of development and future maintenance down. In addition, it provides a common "look-and-feel" for the user experience.

We have developed a configurable query/browse/visualiztion interface library ("Firefly") to allow quick set up of individual mission datasets. Unlike many traditional, web-based applications where most of the processing occurs on the server, IRSAs new web-portal has a heavy client. This offloads the server, takes advantage of client desktop processing power, allows for a smarter, more interactive client, and enhances user experience. The client-side is composed of dynamic HTML, CSS, JavaScript, and AJAX to create a very rich and interactive web application. The client uses Google Web Toolkit (GWT), a development toolkit for building and optimizing complex browser-based applications. GWT compiles Java into JavaScript. This allows us to develop the large application using Java and create reliable JavaScript to run on the browser. No plugins are required to use any of the archives.

To date, this technology is used for the Spitzer, WISE, and Planck archives. In addition we have deployed version 2.0 of the Finder Chart image cutout generator using the same technology. Finder Chart (see Figure 3) allows users to cross compare images from multiple observatories (DSS, SDSS, 2MASS, WISE, IRAS). Although originally intended for generating finder charts for ground-based observing, it has found many other applications. For example, Mathews et al. (2007) screened brown dwarf candidates in r Ophiuchi to eliminate misidentified extended sources. While many implementations of cutout generators exist, Finder Chart takes advantage of IRSAs mission-specific expertise by returning and displaying the 2MASS and WISE images along with lists of image artifacts to assist in deciphering faint sources.

In the next three years, we plan to upgrade the other holdings to provide the same flexible interface. While requiring an initial investment of effort, the long-term cost of maintaining IRSAs web portal will be reduced.
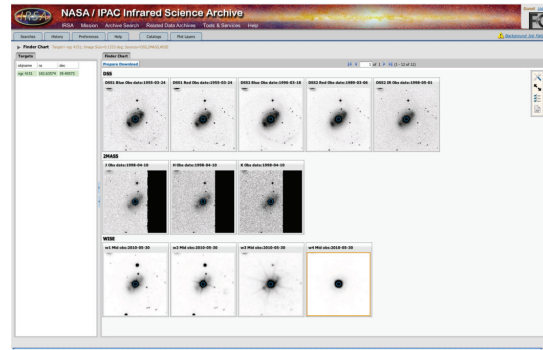
Figure 3.: Screen capture of IRSA's Finder Chart tool, version 2.0 beta, which allow cross-comparison of images from multiple data sets.

4. Future Plans

Going forward, IRSA will be reimplementing much of our core infrastructure in order to support continued expansion of our holdings.

We are in the process of increasing access to the data through the VO protocols. In particular, we are implementing the Simple Image Access Protocol (SIAP) for an increasing number of datasets, with the goal of making all images available. Next, we will develop the infrastructure for support of the Simple Spectrum Access Protocol (SSAP). Finally, we are currently evaluating the feasibility of the Table Access Protocol (TAP) and the Observation data model Louys et al. (2011).

IRSA plans a major upgrade to the Catalog Search and Data Discovery services. Particular emphasis will be placed on the synergy between data sets. Cross-mission visualization of images, spectra, and observational meta-data (such as slit position on the sky) will greatly enhance the user experience.

In summary, IRSA continues a major expansion that is giving us a unique opportunity to improve science return from the archive. With the addition of Spitzer, WISE, Planck and other data sets, we expect to reach almost 700 TB in holdings by 2014. IRSA enables rich and diverse science, with applications across all the strategic areas of NASA astrophysics.

References

Louys, M., et al. 2011, ArXiv e-prints. 1111.2281
Mathews, G., Allers, K., Liu, M., & Magnier, E. 2007, in American Astronomical Society Meeting Abstracts, vol. 39 of Bulletin of the American Astronomical Society, 103.21
Neugebauer, G., et al. 1984, ApJ, 278, L1
Planck Collaboration, et al. 2011, A&A, 536, A1. 1101.2022
Roby, W., et al. 2013, in ADASS XXII, edited by D. Friedel (San Francisco: ASP), vol. 475 of ASP Conf. Ser., 315
Skrutskie, M. F., et al. 2006, AJ, 131, 1163
Teplitz, H. I., et al. 2012, in Astronomical Data Analysis Software and Systems XXI, edited by P. Ballester, D. Egret, & N. P. F. Lorente, vol. 461 of Astronomical Society of the Pacific Conference Series, 605
Werner, M. W., et al. 2004, ApJS, 154, 1. arXiv:astro-ph/0406223
Wright, E. L., et al. 2010, AJ, 140, 1868. 1008.0031