A Case Study in Modernizing Software: The IRAS Scan Processing and Integration Tool (“Scanpi”)

Anastasia Alexov, John C. Good, Iffat R. Khan

Infrared Processing and Analysis Center, California Institute of Technology, Mail Code 100-22, Pasadena, CA 91125, U.S.A.

Abstract. The all-sky far-infrared sky survey performed by the Infrared Astronomical Satellite (IRAS), launched in 1983, remains of exceptional value in astronomy. A tool developed during the IRAS mission, the Scan Processing and Integration Tool (Scanpi) has proven indispensable in maximizing the scientific value of the IRAS data. It performs weighted average fluxes of 1-dimensional (in-scan) IRAS raw survey scans; these averages provide sensitivity gains of 2–5 over the IRSA Point Source Catalog (PSC) in the fluxes of extended, confused or faint sources. It has recently been used in research areas as diverse as searches for planetary debris disks and star formation in low surface brightness galaxies. The aging code, now under maintenance at the NASA/IPAC Infrared Science Archive (IRSA), has proved ever more difficult to maintain and build. Scanpi was written in FORTRAN66, and over the years became unwieldy with the addition of wrappers and patches to keep pace with changing platforms. In 2007, IRSA delivered a modernized version of Scanpi, designed for long term maintenance and offering new functionality. Scanpi was rewritten in C and deployed on a Linux server. A major part of the development was to integrate Scanpi into the IRSA software architecture, which has been in operations for nine years, has supported over 22 million queries and is under active maintenance. Scanpi is written largely in C for performance and maintainability, and supports VO protocols. The architecture is designed as a set of stand-alone and reusable modules with simple program interfaces. Thus existing modules which perform tasks such as coordinate transformations and table filtering have been incorporated into Scanpi. We describe lessons learned and list best practices for modernizing software.

1. What is Scanpi?

Scanpi is an interactive tool for viewing, plotting and averaging the calibrated survey scans from the Infrared Astronomical Satellite (IRAS). Scanpi has proven indispensable in studying extended, confused or faint sources, where it offers a substantial sensitivity gain over IRAS mission products. There is a sensitivity gain of 75% over the IRAS Faint Source Catalog and a factor of 2–5 over the IRAS Point Source Catalog (PSC).

Scanpi remains an important science tool. Between 1993 and 2008, 5 to 15 papers per year cited Scanpi, and the service was accessed up to 25,000 times per year between 2002 and 2008. Most recently, Scanpi has been used to estimate whether the Herschel Space Observatory will be able to measure the

\[ \text{http://scanpiops.ipac.caltech.edu:9000/applications/Scanpi/} \]
host galaxies of quasars. This study (Sergeant et al. 2009) showed that IRAS
100 $\mu$m fluxes at $< 1$ Jy may be overestimated by $\sim 30\%$.

2. Why Modernize it if it Works?

IRSA inherited the maintenance of Scanpi from IRAS. After almost a quarter
of a century the aging code has proven difficult to maintain. Scanpi was written
in FORTRAN66; it consists of a series of multiple and complex layering of
programs, scripts and patches. Over the years, it became unwieldy; it was
difficult to fix bugs, add features, and keep up with platform, hardware and
software upgrades. In 2007, it became orphaned on Solaris 2.8 when it would
not compile on Solaris 2.9, and the data decompression algorithm failed. Lastly,
the Scanpi developer was retiring and we wanted to modernize the software while
the expertise was still in house. IRSA decided to re-write Scanpi from scratch.

3. Start with a Clean Slate

In preparing to modernize and re-write Scanpi, we revisited the original assump-
tions and algorithms; we could improve the algorithms if need be since we had
had many years of experience with Scanpi. Our goals were to:

- Redesign for ease of maintenance, portability and better performance
- Provide more extensive functionality
- Provide transparency inside the processing steps
- Integrate it into the IRSA software architecture:
  - IRSA has been in operation for 9 years with over 22 million queries
  - It has a mature architecture in active maintenance
  - The architecture consists of standalone, reusable software modules
  - The modules create a “Tool Set” consisting of libraries, programs and
    CGI services for generic tasks such as coordinate transformations,
    plotting, table file manipulation and filtering
  - Modular programming tools are easy to plug in and (re-)use
- Provide a maintainable, portable, high performance tool written in C.

4. New and Improved Scanpi

The new version of Scanpi includes IRSA module reuse and algorithm rewrite.
The design of the new version is shown and discussed in Figure 1. In terms of
thousands of lines of code (KLOC), Classic Scanpi had 37 KLOC for the main
program and 65 KLOC for the libraries and plotting. The coding style was
dense, there was no debugging information and very few comments throughout
the code. New Scanpi was written in 21 KLOC with a non-dense coding style
with comments and debugging throughout. The IRSA toolbox contains 200
KLOC, and was used as needed for Scanpi.

In the new version we unpacked the Scanpi input (IRAS level 1 scan) data
into ASCII, human readable, interpretable files. It is much easier to work with
non-binary, uncompressed files. We also added the following new features: ad-
justable fitting and display parameters; all products available from each pro-
Figure 1.: Design Diagram for new Scanpi. A user’s source is read in and we find the IRAS scans which crossed near the user position to create a scan list. This list is processed in a loop where we filter the scans for those within the user’s specified range. We re-bin the data to be on the same scale and perform background fitting followed by source fitting. Individual scan results are stored in the results set. The loop is performed for each scan in the scan list until all the scans are processed. All of the processed scans are co-added and the background and source fitting is performed again on the co-added scans. The co-add results are fed into the same results web pages for the user. This modular design sits on top of the IRSA Tool Box, where generic tools and libraries are re-used as needed.

5. Conclusion: Lessons Learned

We learned a number of lessons from the task of modernizing a 25 year old software application. One important aspect in our case was to store our data in a simple non-compact, non-compressed format. Disk space is cheap while computation is more expensive in terms of cost of CPUs and time to unpack the data. Previously there was too much effort spent to pack data in tightly. In the end, the availability of storage catches up with time and so there was no need to save space.

Another important lesson in writing software to last more than 20 years is not force yourself into a particular, strict software environment. If you use a
component-based architecture, a modular tool set is easier to work with, to fix and to replace entire components. The generic components written for Scanpi, such as the bulk download script mechanisms using wget, are now used by other IRSA applications. The modular architecture benefits the application being written as well as the toolbox. Modules should be written with portability in mind, for long-term maintenance.

Finally, document all your algorithms, and upgrade as needed. In the end, it was not a huge task to rewrite Scanpi because we had the algorithms. In terms of the rewrite level of effort: Classic Scanpi utilized \( \sim 2.5\% - 3.5 \) FTEs + 1.25 FTEs for regular maintenance to keep it running (spread over five years); New Scanpi utilized 1.25 FTEs (including programming, testing and consultancy). In rewriting Scanpi, we refreshed our in-house expertise on this software.

Acknowledgments. We would like to thank Joe Mazzarella and Steve Lord for their assistance and guidance in helping work out the kinks in the new Scanpi as well as bring ideas on how to improve the new version. We would also like to thank Stephen Serjeant for his feedback on the new Scanpi and for helping us work through bug fixes in the software. The Infrared Science Archive (IRSA) is funded by NASA under contract to the Jet Propulsion Laboratory.

References