Astronomical Data Analysis Software and Systems XIII
ASP Conference Series, Vol. 314, 2004
F. Ochsenbein, M. Allen, and D. Egret, eds.

ATLAS - A General Search Service for Complex Data Collections

Anastasia Alexov and John Good

NASA/IPAC Infrared Science Archive, IPAC 100-22, Caltech, Pasadena, CA 91125, Email: aalexov@ipac.caltech.edu, jgc@ipac.caltech.edu

Abstract. Many astronomical datasets consist of diverse and complex data collections (images, catalogs and spectra) of arbitrary size that cover any region on the sky. The purpose of Atlas is to provide uniform access to such collections, using a common interface. We define a set of files to be generated for each collection (image metadata, a list of source tables, *etc.*) and some standards for how data should be organized and referenced and how collection 'Home Pages' should be structured. Atlas is a single CGI-program that can be used to search, subset and present any collection's data to the user via the web. This method of organizing and presenting the data makes it easy to update the data without major modifications or upgrades to the system or software.

Atlas also has a 'mode' to serve image data using the National Virtual Observatory (NVO) Simple Image Access Protocol (SIAP). We will also show how NVO protocols and standards should be extended to support a variety of complex data collections.

Atlas can be found on the NASA/IPAC Infrared Science Archive (IRSA) website:

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

1. Introduction

In future, IRSA will serve complex datasets consisting of set of images, source catalogs and spectra in specific regions of the sky. With the aim of reducing development and maintenance costs, IRSA has developed a single application, Atlas, for querying and accessing these data. Atlas is highly extensible, and supports data sets of any size or complexity. It also supports NVO Simple Image Access Protocol (SIAP).

2. Design

Atlas is a single CGI program written in C. It takes advantage of the component based architecture at IRSA and re-uses many existing modules. These same tools are also heavily used by other IRSA services and applications. There is a substantial amount of code-reuse in this design. The design is data-driven, instead of software-driven. Descriptions of datasets (metadata) are used for

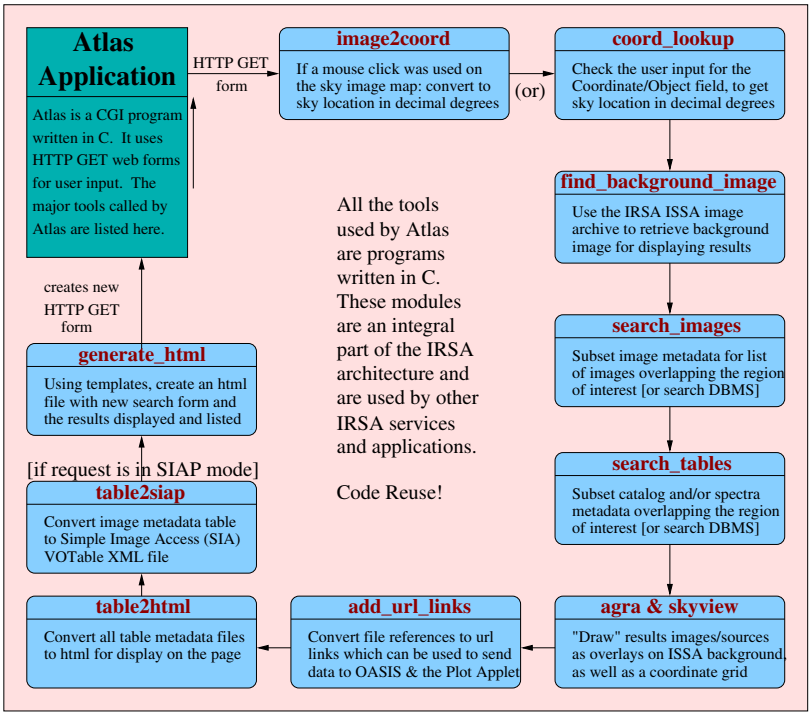

Figure 1. Atlas processing flow

searching data. If a new collection is added or updated in Atlas, we simply (re-)create the metadata instead of changing the software. The Atlas basic processing flow is depicted in Figure 1.

3. User Interface

User interfaces for all the data sets served through Atlas are similar. Each Atlas Data Collection has its own Home Page; this is usually customized to the data providers needs. This page typically has: a sky image with a color overlay of the areas covered by that data set, a data collection description and informational links, and a search form. The user can search for data by either clicking on the image sky map within an area shown to contain data or by entering a coordinate in the "Coordinate/Object" field. The top left hand side of Figure 2 shows the Atlas user interface for one collection.

3.1. Results

Search results are HTML template-driven; the results for all data collections have a similar look and feel. A typical search results page contains "footprints" of resulting data rendered on an a background image of the sky. These query results for all data types are listed in tabular form as a summary of what was found. There are several download and retrieval options for metadata, catalogs, spectra and image data on the results pages. The process of searching is iterative:

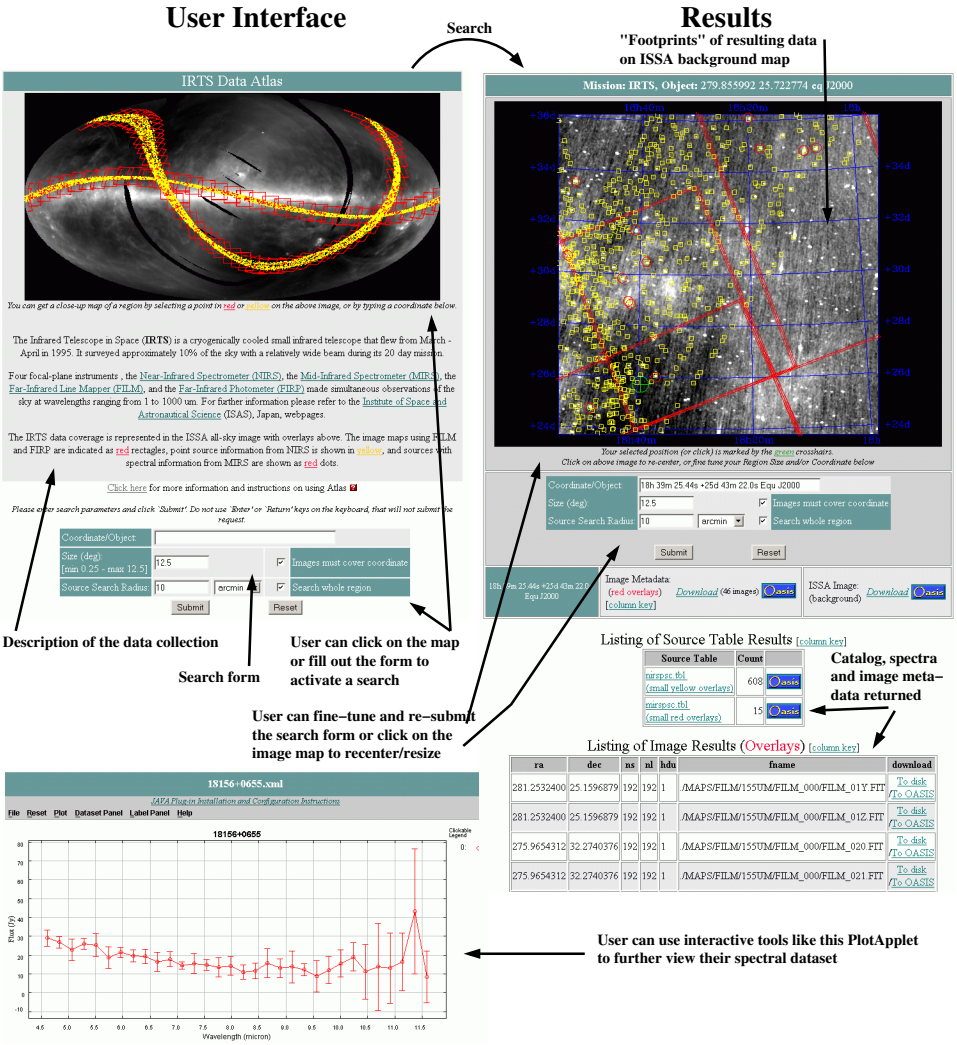

Figure 2. A sample Atlas User Interface for a data collection is shown above. The top left panel is a Home Page which has a visual representation of the data collection coverage, information material and a search form. The panel on the right shows the search results page; data coverage is shown for the request box search of the sky, along with another search form and a list of the data available for download. The user can fine-tune their search, re-size or start over. They can download data and use interactive tools like OASIS and the IRSA plotting applet to further explore the dataset.

the results page is similar to the Home Page where the user started the search – they can fine-tune their search or start over from the results page.

There are many options to download the results. The data collections are served in a number of different formats depending on what was provided to IRSA. Every available format is displayed as a download option on the results page. Some of the download options include:

- Catalogs: ASCII IPAC format
- Spectra: ASCII, GIF, PS, PNG
- Images: FITS, JPG, metadata

IRSA also provides a method of viewing the images, spectra and catalogs using the interactive JAVA tools OASIS (Good *et al.* 2003) and the IRSA PlotApplet.

4. Future Plans

Atlas was released in October 2003 and is being expanded to meet data-provider demands as well as user feedback. The first release serves nine different data collections, which have various combinations of image, catalog and spectral data. The sizes of the datasets vary from collections of only several hundred to over 50,000 sources/spectra/images.

In 2004 IRSA plans on the following upgrades to Atlas:

- Access SIRTF science data sets using NVO SIAP.
- Serve SIRTF Ancillary Legacy Data Collections.
- Add access control (*i.e.* passwords, proprietary periods)
- Add postage-stamp GIF previews of FITS Images
- Implement SQL-like searching capabilities on metadata

5. A Challenge for the VO

This work has revealed the need for extensions to the SIA protocol to support data collections and a broader range of file handles, not just URLs. For instance, Atlas is able to serve a variety of data from one collection; however, only the resultant image data can be served through the VO, using SIAP. We feel that the VO data access methods need to be able to describe data collections (images, spectra and catalogs).

Acknowledgments. We are very grateful to Anastasia Laity for helping us put together the poster for the conference.

References

Good, J. C. *et al.* 2003, in ASP Conf. Ser., Vol. 295, ADASS XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco: ASP), 89