

From: Anastasia Alexov <aalexov@ipac.caltech.edu>
Subject: ADA Systems and ATLAS

To: rhanna@ipac.caltech.edu

title: ATLAS - A General Search Service for Complex Collections of Data (IPAC)
authors: Anastasia Alexov and John Good, IPAC, Caltech

Many of the datasets that will be served through the NASA/IPAC Infrared Science Archive (IRSA) will consist of diverse and complex collections, of arbitrary size, of images, source tables, and spectra that cover specific regions on the sky of arbitrary size. The purpose of ATLAS is to provide a uniform access to such collections. 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 'homepages' should be structured. Then, a single CGI-program can be used to search, subset, and present any collection's data to the user. 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. Most importantly, one program enables us to serve through a common interface multiple data collections, of any size, containing any mixture of images, spectra and/or source catalogs, in a simple and uniform manner.

Much of the tabular data in these collections are of modest size (approximately 50,000 records), and querying and subsetting them are most efficiently performed outside a DBMS. We are therefore developing tools based on Open Source expression evaluators to provide SQL-like tabular search and subsetting capabilities.

We will also show how Virtual Observatory (VO) protocols and standards must be extended to support such complex data collections.