



Creation, Maintenance, and Use of the MAST/STScI Publications Database

K. L. Levay (STScI/CSC) E. Fraser (STScI), J. Lagerstrom (STScI), I. N. Reid (STScI), R. L. White (STScI)



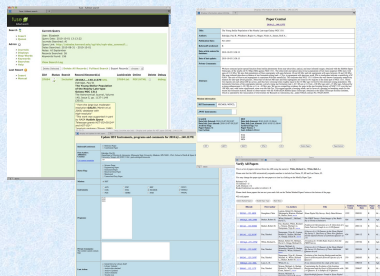
Abstract

Staff from the STScI archives, library and Science Mission Office collaborated to create a database of papers identified as using data archived in the Multi-mission Archive at Space Telescope (MAST) or by members of the STScI staff. This poster will describe the methodology used to identify papers and to populate a database with meta-data about the papers and to link it with the data that are cited within the paper. The database is used to produce a wide variety of metrics for use by the missions and MAST. MAST is developing data discovery methods that utilize VO defined semantic vocabulary to search the abstracts, titles, and keywords of the papers and then utilize the established links to the data to provide an alternative method for mining the archives.

Populating the Database

Identifying Relevant Mission Papers

Curation of a mission bibliography is a time consuming activity so we have automated several steps in the process. To automate the initial identification relevant papers, we use the fulltext search software called Fuse that was developed by Chris Erdmann while at the ESO library. A set of papers to be examined is entered into the interface then the papers are automatically retrieved from the publisher and converted to a text file. A full text search is run to find a list of staff-defined and staff-maintained keywords. Papers with no matches are discarded and candidate papers presented to the staff to determine if the paper should be flagged as mission science paper. We include papers that use HST or MAST data to reach a scientific conclusion. Once a paper has been classified, the papers metadata are retrieved from the ADS and used to populate a database. We then determine the specific datasets and programs and the instruments that were utilized. The staff enters this information via an interface. As part of the software interface, we compare the author names with those names on the original observing programs to determine if the paper is categorized as a GO paper or an archival paper. Some of the metadata change so periodically the database is updated from ADS with any changes (e.g. citation counts).

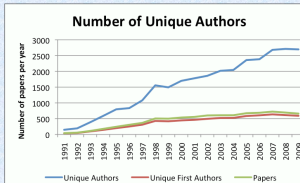
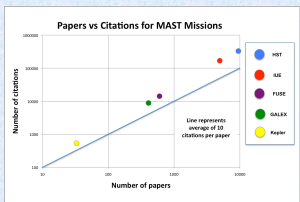
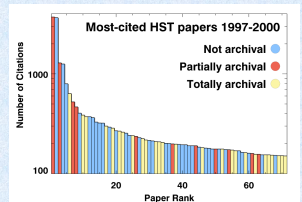
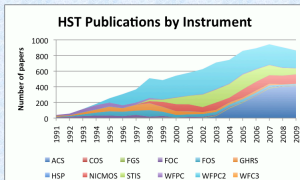
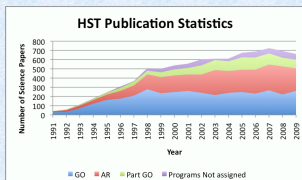


Identifying Relevant Staff Papers

Papers written by STScI staff members are also tracked in the same database, but using different methodology to identify the appropriate papers. An initial software driven ADS search using each staff member's name (and any "aliases") identified possible papers for that person, populated the database. The database entries subsequently verified by the staff person using a web interface. Periodically the search is repeated looking for new papers added since the latest verification. Papers previously deleted are rejected by the software during the updates. If the staff member expects a paper that is not included in the ADS, the librarian works with the ADS to ensure the paper is included.

Metrics

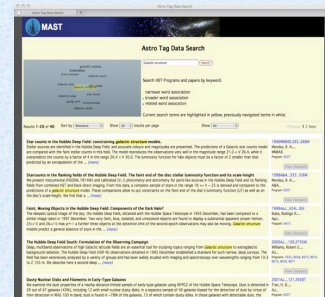
The publication data is used to generate metrics for a variety of purposes. Some of the more common questions start with "How many ...?" as in How many HST papers were published in 2009? We have a web page with metrics for the most common questions about the HST mission (<http://archive.stsci.edu/hst/bibliography/pubstat.html>). Other metrics are requested in an ad hoc manner. By routinely storing values associated with papers such as author affiliation, updated citation counts, and the associated data, a variety of metrics reports can be generated as needed. There are metrics / plots generated to evaluate staff performance that are not shown here. Below are a variety of plots that are have been used by staff members in reports and papers.



Data Discovery Using the Publications

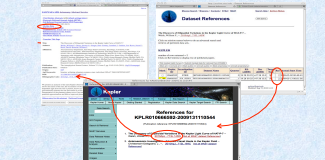
ASTROTAG

MAST has effective search tools to find specific targets or having particular observational characteristics. But it can be difficult to find data on general categories of topics (e.g. strongly lensed arcs in galaxy clusters) because the abstract and target descriptions rarely describe all the interesting characteristics of the objects in a field. Archival users often use data for different purposes than the original GO team planned, and this also includes characterizing the objects differently. MAST has a tool nearing beta release that will allow users to use a selected set of keywords based on the VO semantic vocabulary (IVOAT) to select from and then use to search of titles, abstracts and publisher keywords for papers and programs. A snapshot of such a search is shown to the left. See the paper by Dower for a discussion of the technology being used for this tool.



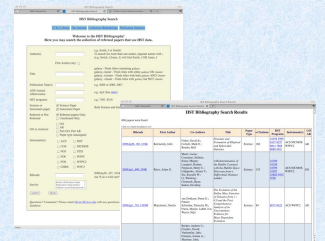
Collaboration with ADS

MAST and the STScI library have a long standing collaboration with the ADS to mutually provide links to the data and the papers. STScI/MAST staff identify the papers using data archived at STScI and provide links to the papers at the ADS from our web pages. ADS also links back to MAST. The MAST interface includes a listing of all papers associated with a specific set of data. ADS and MAST have just begun planning for a more comprehensive collaboration that should permit user to do a more comprehensive multi-archive search utilizing ADS links.



Bibliography Searches

MAST maintains a general bibliographic search interface so users can search for papers for all missions. A search tool for HST bibliography permits users to specify additional types of fields. The search results have links to the papers, the data and where appropriate proposal information.



<http://archive.stsci.edu/hst/bibliography>
<http://archive.stsci.edu/bibliography>