

REPORT OF THE MULTI-MISSION ARCHIVE AT SPACE TELESCOPE USERS GROUP (MUG): FEBRUARY 2005 MEETING

The MAST Users Group (MUG) met at the Space Telescope Science Institute in Baltimore on February 17, 2005. Present were committee members Ann Zabludoff (Chair, U of Arizona), Thomas Ayres (Former Chair, Colorado), James Herald (JHU), Anton Koekemoer (STScI), Jeff Newman (LBL), and Aki Roberge (CIW). Sandhya Rao (U of Pittsburgh) was unable to attend, but sent comments based on her recent MAST use.

This was the fourth meeting since the inception of MUG in 2001 and the first meeting after the successful senior review of MAST in April 2004, in which MAST was ranked second among all NASA data centers (behind only the Astrophysical Data Service, aka the ADS). Many upgrades to the DADS system have improved the typical retrieval times to less than one hour, despite monthly average retrieval rates now in excess of 50 Gb/day and an increase in DADS holdings to over 23 Tb (as of March 1, 2005), including over 400 Gb of GALEX data. The MAST team is to be commended for such great progress on this critical objective.

In this report, we review those outstanding issues that the MUG feels should be addressed in the near term. Generally, the MUG believes that the first priority should be to make MAST data quickly and easily accessible (such as the placing of calibrated data in a quick-access cache). The second priority should be to develop tools to enhance the search and discovery of MAST data. In particular, the MAST identified a strong need for enhancing the VizieR-MAST search tool, searches based on object classifications, and searches requiring the cross-correlation of different MAST catalogs, including the newly acquired GALEX database. The third priority — and only if resources permit — should be to develop simple, easy-to-use tools to combine and analyze datasets “on the fly” (such as Specview). For example, the MUG would like to see smoothing and logarithmic scaling options for previews. In summary, the MUG believes that *to best fulfill its mission, tool development at MAST should focus on improving existing tools and facilitating data retrieval and archive browsing, rather than producing new online data analysis capabilities.*

- **Archive enhancement project.** The HST archive enhancement project is a high priority given the expected acceleration in retrieval times. At the time of the MUG meeting, there was not yet a detailed plan to implement the quick-access cache within one year (nor was this program fully funded). In the interim, the MUG suggested that the cache could be populated by piggy-backing on current DADS extractions. This approach has the advantages of cache'ing the most popular files first and hastening the storage of calibrated data files, while not adding much to the operating load.

Not splitting requests between DADS and on-the-fly calibration (OTFC) is attractive but there are problems to address. A possible disadvantage of this approach is that on-the-fly calibrations (as are currently done for all but the FOC data) are presumably better for being more up to date. The new software could be designed to return only OTFC in cases where the old calibration files are obsolete. In the case of FOC, OTFC is not even possible because some critical files have been lost. The MUG does not want this problem to arise

with STIS data and urges that the files necessary to produce OTFC be obtained and stored now.

- **VizieR-MAST search tool.** The VizieR-MAST search tool is an excellent tool that has significantly enhanced MAST's capabilities. Some major flaws identified by the MUG are really within VizieR itself. Ideally, the MUG would like to see capabilities for 1) filtering out calibration data from the MAST search¹, 2) adding the full set of MAST search constraints (more than just selecting an instrument or mission), 3) returning the results of the VizieR search as well as those of the MAST search, and 4) searching more than one VizieR catalog at once, which is possible with the regular VizieR search form. (This functionality would help surmount the difficulty of doing meaningful searches of VizieR's many overlapping catalogs.)

The MUG had some trouble with VOTable and would like to see more on-line help. For example, at the time of our meeting, clicking on the "Display numeric columns graphically using VOPlot" link at the top of the search results page opened a new window with empty plot axes and generated an error message. Another problem arose when one of us tried to put in a constraint on Dec, which worked when entered as a "User-specified field" but not when entered in the box on the search form labeled "Declination". The MUG member's quick perusal of VOTable also found the "Mark All Missions" button to be broken².

- **The Sloan Digital Sky Survey.** Given the concerns about the current format of the SDSS web interface, the importance of this survey as a complement to MAST's existing databases, and the expertise developed by MAST as the official GALEX site, MUG urges MAST to continue to seek funds to incorporate the SDSS (instead of just mirroring the SDSS site). The MUG would like to see the SDSS photometric and spectroscopic catalogs integrated into MAST. If it is possible to implement this upgrade with the existing MAST interface, then the combination of the HST, GALEX, and SDSS databases would be among the most valuable of astronomical resources. MAST is located one floor up from the SDSS archival group at JHU, which should expedite the upgrade and keep costs relatively low.

- **GALEX spectra.** The MUG noted that there have been numerous complaints about the absence of GALEX spectroscopic data in MAST (and that this has been out of MAST's control). The missing data make it hard for users to propose through the Guest Investigator program. It is recommended that the GALEX users committee, among others, be encouraged to prompt Caltech for these data.³

- **Searches using object classifications.** A recurring theme in MUG reports is the desire for more agile searches by object classification and the cross-correlation of catalogs. For example:

¹This problem now appears to be fixed (June 2005).

²This last problem is now fixed (June 2005).

³We now understand that, as of June 2005, MAST has acquired these spectra.

1) A user wishes to search all HST images for any QSO that lies in back of any cluster (for absorption line studies, etc). At this time, catalogs of all known QSO's and clusters must first be obtained from NED. It is then not obvious that one or both of these catalogs can be uploaded into MAST or that either catalog, once uploaded, could be cross-correlated with the HST archive.

2) A user is interested in all cluster observations within a given search radius of a specified position. It is not obvious how to perform this search with MAST given that the mission catalogs cannot be cross-correlated easily with known clusters. VizieR searches appear to be confined to individual catalogs, not the general class of all "clusters".

The MUG noted that a lot of the hard work in meshing diverse catalogs and setting up object classifications has already been done — in NED. The MUG recommends that MAST explore incorporating the NED classifications to make these kinds of complex searches possible.

- **Searches among MAST catalogs.** Searching among those databases already in MAST can also present some problems. For example, it is difficult for a user to find all objects in the GALEX database with no counterpart in any other database (or, most simply, with no SDSS counterpart over the SDSS volume). This search might be possible using SQL with some finesse, but it would be very helpful to use a simpler MAST interface to obtain the results.

The MUG also recommends that some effort be made to incorporate the footprints of each survey into MAST. Were a user to ask for objects in a region where the catalogs do not overlap, the footprint service would report that back immediately, quickening retrieval times and making possible important science applications such as determining the relative areas or incompletenesses of two surveys. Implementation of such a service would be made challenging by GALEX's existing htm algorithms, but, given that other websites like HESEARC are always being searched by MAST for new or better tools, it may be possible to find an existing footprint service.

- **Some Future Challenges.** How will MAST need to evolve in the near term? MAST is already the official GALEX site and will also be the site for Kepler and JWST. Other potential missions, such as SIM and TPF, are considering MAST as a host. These contracts help fund MAST, provide a homogeneous database across different missions, and create opportunities for data mining and scientific inquiry on scales unimaginable even 10 years ago. But they also present new challenges.

Kepler (launch date: April 2008) will require MAST to handle its first time sequence data. The nature of the Kepler observations will also necessitate an object-oriented archive. For example, Kepler Guest Observers will want their observations immediately, which will be buried in data received every 15m and will need to be separated from other proprietary data. It will be essential to identify those observations 1) of a particular object taken at different times or 2) of all objects taken at a particular time, and to group them together for the user. The MUG recommends that MAST (and the entire ADAC) discuss these new

requirements with other communities, such as solar physics, that are experienced users of time sequence data and big archives. Given that the MAST Kepler archive will include light curves for $\sim 10^5$ solar type stars and a database of ~ 700 planets (with ~ 50 expected to be Earth-like) within five years of its launch, the MAST users community should consider whether resources should be devoted to generally including solar system objects in MAST.

Thanks to Tom Ayers and the rest of the committee for their help with this report. The next MUG meeting is tentatively scheduled for early spring 2006.