MAST RESPONSE TO THE REPORT OF THE MAST USERS GROUP
## OCTOBER 2003 MEETING

We thank the members of the MAST users group (MUG) for their recommendations. We are submitting this response to update the MUG on how we are responding to these recommendations.

DADS upgrade: As noted in the MUG report, last year's MUG meeting took place shortly before major upgrades to the DADS system. The new DADS 10.2 software was installed on December 3, 2003. We are pleased to report that the upgrades have been extremely sucessful. Retrieval times have been dramatically reduced, and many new capabilities have been provided to users, including several that were mentioned in the MUG report. Some of these improvements include: availability of secure FTP for "pushes" of proprietary data to a user's machine and password-protected staging and ftp retrieval of proprietary data, the ability to request only files with certain extensions or raw data only, and improved error messages and status reporting.

Retrieval times: The MUG report recommends that MAST take all steps possible to ensure the fastest retrieval times for its users, as well as to anticipate potential gridlocks in the system such as associated with major new data releases, proposal deadlines, etc. We agree that ensuring that our users can retrieve data in a timely manner should be our highest priority, and we are pleased to report that the installation of the new hardware and software described at the meeting last Fall, as well as increased bandwidth, has indeed resulted in dramatically improved retrieval times as anticipated.   The new EMC "spinning disk" array allows much faster file retrieval and is far less subject to mechanical difficulties than the old "jukebox" system. Currently, although most of the data are retrieved from the spinning disk array, the software is still coupled to the jukeboxes, so if a jukebox goes down, it could still bring the entire system down. We are working on changing the software to decouple the jukeboxes completely from the retrieval system. Another factor speeding up retrievals is that On The Fly Reprocessing (OTFR) is now done on the powerful multi-CPU SunFire machine, which speeds up the processing enormously. Finally, our bandwidth has been increased to 25 Mbits/sec (up from the 11 Mbits/sec that it was restricted to in the August/September time frame), and we can now push data through Internet 2 (50 Mbits/sec) for those institutions that are Internet 2 subscribers.

For the past four months, median retrieval times have typically been under 2 hours, and have frequently been as short as 30 minutes (see chart below). Retrieval times have remained reasonable through several potentially challenging periods, such as the HST and Spitzer proposal periods and the release of the Ultra Deep Field data. We have also adopted several further policies, along the lines of the MUG recommendations, to ensure that we keep our retrieval times as short as possible:

- Anticipate times of high-traffic activity, such as proposal deadlines, and avoid scheduled downtimes during these periods.

- Anticipate release of future large volume or popular datasets (such as GOODS or the UDF) and pre-emptively set up mirror sites and ftp sites so that requests do not

Retrieval Times since DADS 10.2 install 12/3/03

Past Year's Request Retrieval Times
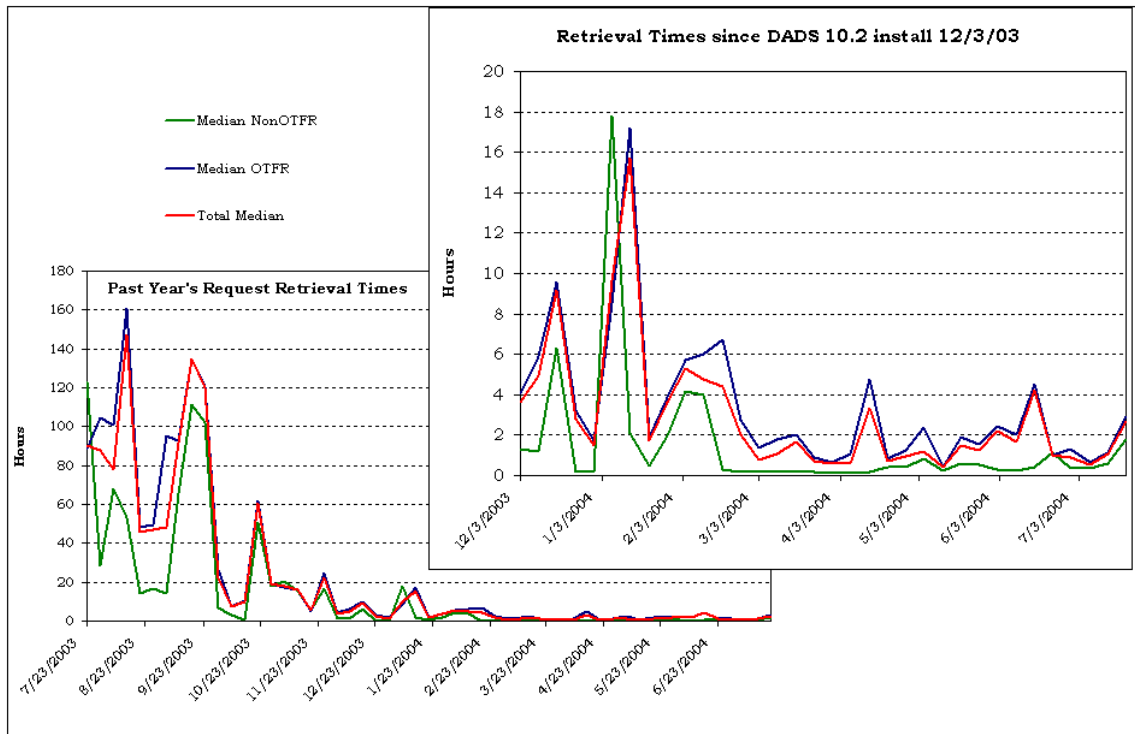
have to go through DADS.

Retrieval priorities: The MUG report recommends that MAST adopt a priority system for retrievals, in which proprietary data and small requests are given higher priority. In the old DADS system, retrievals proceeded through a single queue on a first-come, first-served basis. This contributed to delays because bottlenecks could occur behind large or problematic requests. In the new DADS system, archive operators can change the priority of requests manually or by user id to prevent these bottlenecks. Moreover, because of the much faster processing times enabled by the new hardware, such bottlenecks simply have not been occuring as they were in the past. Combined with the very reasonable retrieval times that have become the norm, we feel that this removes the need to implement a static priority system based on request size or proprietary status, as recommended by the MUG.

Quick-Look FUSE data: The MUG report recommends that MAST make FUSE wavelength-flux data available online for direct download as an ascii or fits file, so that quick reconnaissance of a data set can be done without submitting a DADS request. MAST currently provides some sort of 'quick-look' preview plot for all but two FUSE observations (for which the FUSE project has not provided us with the preview data). For the data that have been through the processing pipeline which was implemented in September 2002, we also have more comprehensive preview plots (provided by the FUSE project) with a link to the interactive plot option. (As of July 27, 2004, 1982 of the 3615 FUSE observations had been processed through this pipeline. Of these, 1013 are still proprietary, so currently 969 observations have the enhanced preview option available.)   This past Spring, MAST scientist Myron Smith communicated with the FUSE project scientist, and was told that given the improved DADS performance, the

project did not see a need to pursue the ascii data download option at this time. As well, MAST has been actively considering other ways to streamline access to the FUSE data, such as migrating the data from DADS to the EMC storage system so that all the data would be directly accessible online. We are continuing to pursue this option for the long term because of the enhanced data reconnaissance and Virtual Observatory applications it would enable.

Object Classifications: The MUG report expressed the need to enable searches of the MAST holdings by object type. As noted, the HST key words are inconsistent and highly unreliable for this purpose. The MUG members are certainly aware of the difficulty of assigning object types to large and diverse data holdings such as those hosted at MAST. As a partial solution to this problem, we have recently introduced the capability to search the more than 4000 catalogs hosted by the VizieR service through the MAST portal (http://archive.stsci.edu/vizier.php) and to then cross-correlate catalog entries with any MAST mission. Users can search for catalogs based on catalog name, mission, target name, wavelength range, keyword (author name, title, description), or object class. Users can now also upload their own catalogs and cross-correlate them with MAST's holdings. We will continue to develop further capabilities to enable object class based searches, in accordance with the MUG's recommendation, e.g. by accessing object type information from NED or SIMBAD.

Sloan Digital Sky Survey: The MUG report recommends that MAST extend the GALEX interface to allow simultaneous searches of the SDSS and GALEX archives, as well as MAST's other holdings. We share the MUG's wish that we incorporate the valuable SDSS data with the MAST archive, but our current MAST contract does not permit us to devote our resources to SDSS. However, we are pursuing other options – we are currently writing a proposal to NASA for significant enhancements to the HST archive and plan to include in that an option to integrate the Sloan data (which would clearly be a very useful addition at a modest cost.)

StarView: The MUG recommends that MAST should pursue migration of StarView functionality to the web interface, and devote resource priority in this direction rather than to upgrading StarView. We note that currently there are essentially zero MAST resources devoted to maintaining and upgrading StarView. We are working on implementing through the web interface specific capabilities that users and the MUG have expressed interest in, such as aperture and object/image overlays and more powerful spectral preview tools.

MAST again thanks the MUG members for their helpful comments and recommendations.