MAST Users Group Report – November 2012

Version date 31 Jan 2013

The MAST users group (MUG) met at the Space Telescope Science Institute on 26 November 2012. MUG members are Megan Donahue (Michigan State University; chair), Jay Anderson (STScI), Ed Fitzpatrick (Villanova), Suvi Gezari (University of Maryland), Timothy Hamilton (Shawnee University), and Guy Stringfellow (University of Colorado).

We listened to powerpoint presentations from MAST and STScI personnel, all of which are also available on the MAST website. To the extent possible, we regard this report as a response to those presentations, and we will not attempt to reproduce the content of those presentations here.

After the close of this meeting, Jay Anderson and Megan Donahue rotated off the committee, although both members participated in the preparation of this report; Donahue collated and edited the final versions. Guy Stringfellow was chosen by the remaining members to be the MUG chair of the 2013.

Overall Comments

The Mikulski Archive for Space Telescopes has enjoyed a productive year both in science and in the development of useful tools and capabilities for astronomy. In particular, data from the Hubble Space Telescope, Kepler, and GALEX are retrieved at high volumes by many users. The High Level Data Products from the Hubble Legacy Archive and legacy programs such as the Hubble Multi-Cycle Treasury programs are extremely popular, as their retrieval data volumes exceed archived (stored) volumes by an order of magnitude. The infrastructure upgrades to the servers have paid off in noticeably faster retrievals. Furthermore, this is a global service, with data recipients all over the world (all continents except for Antarctica were represented.)

In the following sections, we have endeavored to identify areas of opportunity and of concern, and, where feasible, we have made some suggestions to MAST for prioritizing work. In all cases, our main interest is to maximize the scientific productivity of researchers using the MAST data and services.

User Survey 2012

The results of the 2012 MAST users survey were presented by Anton Koekemoer. Approximately 300 responses were received out of ~6000 requests; (a ~5% response rate).

There was little discussion of the results of the survey but a number of important points were evident:

- Most respondents listed themselves as using the system either a few times per month (110 out of 307) or a few times per year (122 out of 307). The predominance of occasional users suggests that many users will not be experts and so the MAST interfaces should be as intuitive as possible, allowing basic access and basic data retrieval with a shallow learning curve. The availability of deeper functionality, for more expert users or more complex searches, should be apparent from the top page.

- Users were generally satisfied with the current performance of the system, but a number of "wish list" items showed up that the MUG also agreed with: more (and easier) access to MAST documentation of current services and the most recent instrument documentation; as many cross links to as many other archives/databases as possible (e.g., 2MASS); cross correlation services with other missions; integration of all MAST services to a single site; an increased number of tools for manipulating and filtering tabular results and other data after identification.

- A number of users commented that the current interface is not intuitive. Some effort might be expended into interviews and targeted user-surveys with users to try to identify what "intuitive" might mean. The MUG understands that there is a natural tension between the user's desire for simplicity and (sometimes the same) user's desire for many features. As noted above, a site with a simple entry point that allows straightforward searches but points the way to more high level, highly featured use of the archive might satisfy all these desires.

- One of the recurring help-desk and survey comments involves providing real-time feedback to MAST users when data retrieval is taking longer than normal. Providing system stats and live retrieval rates is good. We are aware of the ST-DADS status page; it would probably be a good idea to figure out why that status page is not known (users don't know where to click to find system status) or why the information on the page is insufficient. We suspect that users might appreciate what some other systems provide is a "real-time" update of the status and typical response time over the last hours, not just averaged over the last 4 days.

We noted that Kepler was one of the most heavily (and largest) downloads, yet only 11% of the survey participants identified accessing Kepler data. We were not sure whether largely automatic downloads meant there was a large load placed on MAST derived from a small segment of the community.

In discussion, MUG members had concerns about the survey instrument itself. In particular:
- the low rate of response (are we getting a representative subset of users?);
- the breadth of the questions being asked of all respondents (e.g., not all users could knowledgeably respond to questions about the Kepler database)
- the questions/answers (e.g., several interpretations could be placed on the survey answer "Not Sure")

Recommendations for future MAST surveys:
1. Consider conducting targeted, shorter surveys that would only take a few minutes to complete. Such surveys or questionnaires could be aimed at users of specific datasets (e.g. Kepler) or of specific interfaces (e.g. HLA). Since the return might be sporadic, the occasional open-ended question, posed as they leave the page, such as “Did you find what you were looking for? If not, what was it?” might generate some interesting responses.

2. For the next overall user survey, make sure that choices for a given question are relatively unambiguous in their interpretation. Bonnie Eisenhamer in OPO has provided advice and assistance in previous institute surveys, and there are likely other experts on staff that MAST could ask for advice.

3. Yield and representation for returned surveys are perennial issues. Consider experimenting with additional strategies to maximize yield and ensure broad representation of your diverse user base.

4. It might be interesting to see if there are differences of opinions when responders are split between internal/external or experienced/inexperienced users. Give some thought about how to define veteran/novice or frequent/infrequent, and how those user groups might differ in their experience of your archives. (You have veteran external users who use MAST infrequently, but who have been retrieving data for over a decade, for example.)

5. You might consider making some questions contingent on whether the user expressed experience with a given service. For example a modern marketing survey will often ask a customer what services they are familiar with, but then drills down to asking questions only about one or two of those services, or asking questions only about the services the respondent indicates they knew.

6. In the interest of user-focused development, maybe the next MUG meeting could allocate time for watching individuals attempt to use your interfaces to accomplish a task. You could set up one-on-one sessions, for example. Another source of beta-users or novice users could come from JHU astronomy graduate students (and pay them in free food, gift certificates, or other incentives.)

In summary: the MAST survey is seen as a good source of ideas and a reasonable reflection of user satisfaction and familiarity (or lack thereof) with MAST, but it could be even better. The MUG agreed with many of the suggestions that emerged from the survey, and encourage MAST to consider these suggestions when prioritizing their efforts for the coming year.

Common Observation Archive Model

The goal of the CAOM is to unify the archive, consolidate resources, and provide a single, unified interface to interact with data from all current and future missions in ONE place in the

SAME way. The CAOM is VO compliant, allowing for CAOM data to be broadcast to VO tools. The CAOM also allows for more flexible searches of data, across multiple wavelengths, since all data from the different missions are visible.

MAST is collaborating with the CADC to develop CAOM v2, an enhanced version of CAOM v1, which will have more features specific to the MAST archive. Having only one unified format makes maintenance of the archive much easier.

The conversion to the CAOM format is being done outside of the data archive catalog. This strategy leaves the mission databases unchanged, and allows for a composite CAOM database for science queries. The current databases will continue to be populated, and then later converted to CAOM. In the future, MAST plans to make population of the CAOM database to be a part of the pipeline.

The MUG notes that there have been some issues with high-level data products provided by the user community that do not have enough metadata to conform to CAOM format. The MAST MUG encourages MAST to participate early and vocally in establishing and communicating the data requirements so that new missions and high-level data products conform to CAOM. Communicating with missions early in the process, providing assistance in the form of education and code libraries, will alleviate the load on MAST resources later to convert data to CAOM compliance while minimizing the cost to the developing mission. MAST representatives can participate in the review of missions' data management plans, for example, if MAST is being identified as the data host for the mission or the archive.

Adopting CAOM is of benefit to all data archive centers. The more common the data model, the more easily data is shared across archives. Furthermore, there is likely to be cost savings to the mission if they are adapting to an existing template as opposed to inventing their own from scratch. The MUG suggests that MAST continue to actively cultivate collaborations, such as joint projects with other archives. It is useful to avoid the "not invented here" attitudes that occasionally form in these interactions, historically. The MUG was encouraged that many of the standards being set up in the CAOM are similar and possibly compatible with standards from the CADC's archive.

Data Discovery Portal

The Data Discovery Portal will be the new website for archive access; a single user interface with the HLA and GalexView as models. It will allow access to non-MAST missions, as well. The MUG was in general enthusiastic about the concept of the data discovery portal, but it did not have a clear concept of the actual details. For example, a portal could have several interfaces, one for the most commonly-requested services, and another for more sophisticated queries. A portal could have a high level GUI that would satisfy 90% of the needs of the users but also could allow (via another interface) for SQL-based queries with a supply of templates and example use cases. The user community has increasing familiarity with catalogs such as SDSS and 2MASS, as well as NASA data archives.

MUG members had some specific detailed suggestions which we capture here. It may be too early to be drilling down to this level, but after the data portal demo, we felt it might be useful to hear suggestions at this level of detail, since it's not easy to pull in external users at early stages of development.

- In AstroView, the products selection requires clicking on the **edge** of the footprint, but some users might prefer to select every product that **includes** the location on the sky you click on.

- For the uninitiated, please provide explanation for the DAOPHOT and SExtractor catalogs, including the distinction between them, in documentation and tool tips (where feasible).

- We recommend that users always have (optional) access to individual exposure (FLT) files in addition to the drizzled files (DRZ).

-We recommend implementation of succinct "tool-tips" with instructions and/or information pertinent to the task at hand. For example, when preparing a request for data retrieval, when a user selects a particular instrument, having a popup window appear remind them of the relevant file extensions for that specific instrument. Options could include "Recommended minimal analysis set: extensions xxx", "Recommended for manual mosaic assembly: extensions xxx", and "Recommended for full reprocessing: extensions xxx" categories would cover most retrieval intentions. This effort would hopefully alleviate excessive (and unintentional) retrieval of all files pertaining to a particular data set when in fact perhaps only one or two files are really needed. Implement a cautionary warning not to use the back-arrow in the browser window if that would reissue the data retrieval request.

- An interface with a dizzying array of different features requires more documentation to let users know what they **can** do with the archive. To make the most effective use of the archive, users must be educated not just how to do the tasks they know they want to do, but as to what tasks are even possible. There are two levels of education: top level is a quick set of options that are intuitive, enabling quick and efficient maneuvering and sensible data retrieval; links to more extensive manuals, webinars (video instructional demonstrations), and the like prepare the user for more advanced use.

- Mouse-over suggestions ("tool tips") and short tutorial videos would be good ways of explaining features or providing a view to the meta-data in a catalog.

-The heaviest or most experienced users might expect the Data Discovery Portal to do certain things they can already do with other interfaces (trivial example: convert between linear and log plots). All of the capabilities of existing MAST interfaces might be expected (but perhaps unnecessary: some triaging and prioritization might allow MAST to refine the scope and produce a useful if not fully-featured portal more quickly). At least the primary interface - the first thing a visitor sees - must be easy to use to conduct common searches and submit requests.

It will be nice to see a working version of this portal soon; the MUG members are willing to make previews of beta versions.

Hubble Legacy Archive

The HLA has been providing a visual way to explore the archive for several years now. Not only has it provided direct and immediate access to high-level products, but it has been working to push the envelope in terms of making these high-level products better.

Recently, it has shown that improvements can be made in stacking images, both through sky-matching and improved astrometry. It is good that they have focused on getting real products out for WFPC2 and ACS/WFC before branching out to other detectors. The additional data products that it provides, such as depth maps, are also useful to the community. It is steadily increasing the number of datasets that have optimized high-level products in the HLA, while at the same time continuing to provide visual access to the more standard MAST association-level products. The HLA has also taken the lead in helping users visualize the Hubble source catalog, as well as other available catalogs, in the context of HST observations.

The HLA is beginning to tackle the complicated task of improving the absolute astrometry of its products as well as the inter-image registration. Absolute registration is important, but it can be challenging particularly when external absolute catalogs are spotty and imperfect. We hope the HLA will push this forward at a reasonable pace, so as to take advantage of available catalogs but not spend too much time on calibrations when the available calibrators are sparse. In the past, there were occasional gross (>arcsec) errors in HST astrometry, many of which were repaired when the HST Guide Star positions were updated with those from a better catalog. However, MAST should be sure that additional effort made to improve astrometry is scientifically justified. Absolute astrometry requires some reference frame, and the best reference frame for the science may depend on the scientific usage of the data. For the best science in all cases, the HLA should provide clear documentation about its default reference frame and the systematic uncertainty for each observation and instrument.

Past MAST-user comments have been that there needs to be some way of including high-level spectral reductions in the archive. The effort the HLA has been making in regards to spectra is laudable. The main archive is not formally tasked with this, but it is clear that the community would benefit from visualization and combination tools, and the HLA is in a good position to get feedback from the community (for example, from the conference that it recently ran at STScI) and to provide best-effort products and develop useful visualization tools. We expect that, as with the image analysis above, these products will soon find their way into the more standard archive products.

The innovations that the HLA has been making should not stop with the HLA. It would be useful to find some way to return these improvements (such as improved astrometric accuracy and

sky-level matching) to the archive so that the pipeline or custom-run reductions could take advantage of them. Headerlets are one way of doing this, and maintaining an archive of them may make sense. But it seems awkward to have relevant information in different places and not attached directly to the fits files themselves. It would be easier to have these improvements be attached in some way with the _flt products in the archive, either as additional high-level keywords in the standard header or as headerlets within the _flt file. Information already does flow back to FLTs from the pipeline's Astrodrizzle stage regarding CR-infected pixels, so it should not break the paradigm too much to allow additional high level information flow back about sky backgrounds and astrometry.

On a related note, many things are changing and we believe the new portal may allow for things that are currently disparate to be unified eventually. For instance, the HLA can list individual FLT images but not serve them up, and the archive serves up FLT images but does not list them. When the HLA starts serving up FLT images, they will have different (improved) header keywords than those in the archive. We understand the need for these differences now, but efforts should be made to converge on a smaller, more uniform set of products. We expect the migration towards a more static archive will make this easier.

Hubble Source Catalog

This project plans to make a master list of sources from all (or most) HST observations. The anticipated catalog limit for inclusion in the catalog is set to a 10-sigma detection threshold in a single visit. [We note that the MUG was not 100% certain about the exact definition of the threshold; it might be worth repeating at the next meeting. 10-sigma sounds high for a formal detection threshold.] For reference for future MUG members, an HST "visit" is defined as the data collected within a span of fewer than about 5 orbits, determined by a single set of guide stars. (A new "visit" might happen during a longer sequence, when a new set of guide stars would be required.) The expected use of this list includes data discovery (finding fields with HST data), cross comparison to independently-generated catalogs, and cross-referencing to other missions (SDSS, PanStarrs, LSST, 2MASS, etc.) The expected level of direct scientific use at this time is being conservatively estimated at perhaps 10-30% of users. The HST source cross-matching feature was the biggest "desire" from user discussions. Along with identified astronomical sources, a side benefit of these cross checks will be much improved astrometry. There was a NASA AISRP grant (end date June 2012, PIs: Lubow & Budavari) that produced an initial catalog from ACS/WFC and WFPC2.

The new science that will be enabled by such a catalog includes searches for sources with time variability and studies of large, uniformly reduced and measured collections of datasets. assembling science working groups

To fully access the data spanned by HST programs, users will need and require advanced query abilities. To facilitate this type of access, an SQL CASJobs type of interface is envisioned, but perhaps looking at ways of helping users to generate such a query, via a web-form for example, to generate templates would be productive. Having a substantial library of templates

has been useful to other programs (e.g. SDSS), so that would be a minimal support item to allow users to construct their own queries.

Value-added, scientifically “clean” catalogs, optionally with external properties (redshifts, cross-IDs) -- like SDSS sub-catalogs of quasars or galaxies -- would be very useful.
A beta version of this catalog has been on-line since June.

MUG Response:

The HLA should also take care that the catalog provide calibrated magnitudes, at least for the point sources. This calibration should be consistent with other calibrations that have been done for the high-level catalogs, such as the UDF. Some iteration with the instrument teams may be required for this.

Caution and thought needs to be exercised in constructing the Hubble Source Catalogue, particularly in the area of "Band Merging". Some sources appear in or are absent in different filters/wavelengths. Reliable astrometry also is a factor. One well known virtue of HST is the high angular resolution provided. Sources will be revealed on the highest angular scales. It will be important to understand how source blending and the above referenced band-merging will be dealt with, particularly if HST astrometry could differ by 0.2" or more between fields observed in either different filters or at different epochs. Cross-referencing to other catalogues (e.g., SDSS, CXO, etc.) poses further complications, especially since many of these suffer from much poorer angular resolution. These questions aren’t new for the cross-identification community, but users should be made aware of the caveats and the possible failings of such cross-identifications.

Those leading the HSC efforts must involve discussions with the community and other cataloging teams to consistently address these issues. Not addressing these concerns could result in spurious identifications and source properties being pervasive throughout the literature.

Any improvements implementing more secure astrometric solutions and the like should somehow be fed back into the general archival data, available to the community; all available data in the archive and reflected in the HSC should be consistent. One possibility to consider might be two HSCs, one for extended sources and one for point sources. Each source list would have its own set of caveats during source selection. [Analogous comments on aspects discussed here can also be found in CAOM and HLA sections above.]

There is some value in considering the HSC as a Hubble Source List instead. A catalog is usually constructed for a specific scientific purpose. Limiting the HSC to sources with greater than 10-sigma detections will also mean that identifying the sources in other frames (with the same 10-sigma threshold?) will be less problematic, especially for point sources. Again, a science working group with some external participation would be extremely useful if it is not in place already.

The MUG regards the Hubble Source Catalog project as potentially very useful scientifically. There are enough choices that must be made about astrometry, source selection, and source measurements that a scientific working group is recommended.

In this presentation and in the presentations on the HLA and the Data Discovery Portal, the MUG heard a call for beta-testers and working group volunteers. We encourage MAST to invite MUG members as well as frequent users of MAST services to review working design documents and even test best-versions of the interfaces. Another useful group to recruit for this type of activity would be relatively new astronomy graduate students, who could bring fresh eyes to user interface testing, and who would represent the most common new user: the junior graduate student.

HST Data Processing and CRDS (Calibration Reference Data System)

The hardware/software upgrades currently being implemented to the HST data processing and archive storage are a welcome step to enhancing MAST data retrieval. Storage of the processed HST data into the Archive DB and/or Online Archive is a big step forward.

The efforts to properly account and correct for the WFC3 persistence issue (presented in the Overview) is important, and provides a significant improvement in the overall quality and reliability of the WFC3 archival data. These efforts need to be incorporated into the overall archiving of all WFC3 data into the permanent archive. Performing these corrections are complicated and convoluted (involving a time history of imaging frames spanning multiple PI programs), and something that the community could not effectively and efficiently do themselves, and certainly not soon after their data are obtained (as access to the necessary images are proprietary). The MUG has a very strong recommendation on this effort: *As these corrections are so essential to the overall science value of the archive and in the integrity of the data, we strongly encourage the MAST team to find a means to have these corrections applied uniformly across all data entering the MAST archives. To this end we encourage cooperation with and involvement from the instrument team responsible for the calibration of WFC3 data.*

Astrodrizzle was not described clearly enough to the MUG for us to provide any productive reactions. As we prepared this report, we did not understand how Astrodrizzle differs from previous drizzling techniques, what elements of enhancement it adds, and what problems might be encountered in its implementation. This topic would be useful to discuss at the next MUG meeting. This is also an example of where MAST needs to educate and inform the community upon release.

Some emphasis was made on the new reference files, and the portability of the processing. High-energy missions (e.g., CXO, XMM-Newton) use such a scheme, but users of these data often need to reprocess their time-tagged data. Reprocessing is not typically required (or desired) by many HST archive users, since updates to calibration files typically affect only the most recent (and usually still proprietary) data. MAST processing of HST imaging and

spectroscopic data should be viewed as accommodating most of the community user needs through direct access to processed archival data. On-line, rapid access to data is important in considering accessing these data in the era of impending time-domain observing, where rapid access to high-quality archival data is needed to evaluate the nature of transients. Reprocessing archival data using the most current reference files and re-archiving should be a priority and routine procedure in the future with the new BAR/WFM/CRDS.

JWST DMS (Data Management System)

JWST's Data Management System differs from HST's, requiring more information regarding telemetry, such as the wavefront sensors, and an order-of-magnitude increase in computational and I/O throughput. The Launch Readiness Date of 2018 allows time to implement these changes.

The need to incorporate so much data from the wavefront sensor is going to put heavy requirements on the JWST part of the archive, but it's unclear how the user will need to use the wavefront sensor data. It seems that some kind of analysis of WFS information may be required to get the best use out of JWST data, but this will be a new mode of data analysis for most users. It would be useful to know whether it will be invisible—a black-box processing step—or whether it will require user choice and intervention.

Another feature that will be new to most users will be JWST's integral field spectrograph. Being both imager and spectrograph, will need a very different way of being incorporated into the search and display page. The new DS9 software's ability to display this kind of data could give some ideas, and Tom Donaldson says that they are looking at a way of incorporating it.

GALEX

NASA operations of GALEX completed in early 2012; mission currently being operated & funded by an international consortium of universities.

MAST will shortly have received (by the end of 2012) all the NASA-supported observations, particularly the ultimate source catalog, the "GCAT."

Since the FUV sources in the new GCAT will be extracted from positions based on NUV images, some FUV sources, which don't have NUV counterparts will be left out of the final GCAT. Myron notes that the FUV-only sources will still be available from the GR-6 and GR-7 data releases. The MUG suggests that it might be wise to incorporate a FUV-only source list in the final GCAT.

It was a concern to the MUG that the exact arrangement about the public nature of data taken during the post-NASA era was not known. According to the press release, the data are public, but the mechanism and the delivery interval were not defined.

MAST's major GALEX-related activity in the near future will be the creation of the "Photon List" files (x,y,t storage for all received photons!). Flux-calibrated light curves will be produced, along with preprocessed light curves, movies, etc. All of these products will be available to the public via MAST. The effort required of MAST is not very high. Months of computer processing time will be required, but MAST staff told us it was not a major personnel effort in terms of human time.

The highest priority MUG request: *Availability of GALEX data and quality assurance needs to be ascertained for the post-NASA period. This action may require interceding by the Program Scientist and Program Executive at NASA HQ.*

Kepler

In past MAST meetings, it was clear that the data that were delivered to MAST from Kepler were not in good shape for public dissemination. MAST seems to have done a good job iterating with the Kepler team to put things into a useful format for the community. They now provide a map of the field, in addition to a variety of catalogs that can be used for searching, both for objects currently in the archive and for objects that could be proposed for. They have also made many of the calibration products available.

One of the remaining issues about the quality of the Kepler data is that the Kepler Science Center focuses on extracting very localized differential photometry, and tracking down all the issues that affect photometry over the very local time period relevant to planet detections. Not much effort has been extended to treat the variability over larger time periods and between quarters. Given that much of the community is now using the data to explore variability on larger timescales, it would be good to find some way to improve the long-term normalization, either through the Kepler team, through an enterprising archive proposal, or through MAST's own efforts. It seems wasteful to have every user develop his or her own strategy for this. Even though the baseline matching probably cannot be done with the same accuracy as the differential photometry, it would be good to have some kind of best-effort normalization. For many applications, best-effort is all the users need, as the full Kepler precision is not needed to explore many of the long-term-trend phenomena. MAST should channel the interests of the community and press for a solution to this, as they have on several other fronts already.

The idea of letting users download the entire set of data for one quarter was a nice way to get a lot of data to the community and allow them to conduct searches in an automated way. The archive is also exploring useful tools to allow some kind of fit-related searches to be done through the MAST portal.

MUG suggestion: *Matching of the photometric light curves between monthly and/or quarter periods for all sources would provide a high return to both the community and science, if done sooner than later. It would be beneficial if the Kepler team, who has the expertise to enable this*

to be accomplished, could work in a timely fashion with the MAST Team to enable this now. This mismatch of photometry being provided by Kepler to the community through MAST can not be afforded the distinction of high quality if light curves are mismatched between adjacent time intervals.

Big Data and The Future

This topic will deserve more scrutiny at the next MUG meeting, as it is a new effort and little time at the end of the day (with time expired) allowed much discussion. This is an area that deserves liaison with other efforts and teams, most notably the SDSS and LSST teams. There is much, both past and future, to be gained in these efforts with extensive interaction of MAST with these other teams, and with the community at large.

A notable comment in the presentation was how to leverage expertise within STScI in dealing with large data sets. There are “Large Data Sets” already residing and being accumulated by HST, which should not be overlooked or diminished within these efforts while looking to the future. There are several Legacy and Treasury programs producing quite large data sets which could benefit from large-scale mosaicing and uniformity issues that the team could be pursuing internally.

While the motivation for providing SWIFT UVOT and XMM-Newton OM data within the MAST archive is understood, provisions are needed to assure that all data accessed through MAST mirrors that available at the time through the primary archival homesite (HEASARC). The danger here is that community usage through MAST undoubtedly assumes they are getting the best and most current data, which if retrieved from MAST is not currently guaranteed. Inclusion of these other archival data sets into MAST should be performed only if, through working arrangements with the primary archival homesites, it can be guaranteed within some minimal time period that the MAST archive is updated as required, and thus reflects the data holdings in the primary site; MAST needs to be a current mirror site of all holdings available within the primary archival site.

The effort to provide a direct link from the data records to the literature using that data is a useful scientific effort. However, the details of the effort, and the degree to which each literature item is apparently scrutinized caused us concern initially that much effort was being made to identify the degree of usage (primary/secondary/not actually used) and to make fine distinctions between "PI" use and archival use. It may be possible to do these detailed investigations over a representative sampling of the MAST literature. Trying to track down fine details in individual papers may be more work than it is worth, scientifically. Much of this likely can be handled by more direct, effective communication with the community users as to what information MAST requires. Upon publication users could upload bibliographic information directly into a MAST site, ready to be assimilated by MAST. We (the MUG, MAST) should continue to lobby the broader community for better reporting on data usage, from the journals to the interim and final reports going to the funding agencies, and for the results of that reporting to close the loop with the data providers. Action item: *We recommend regular reviews to report*

progress in coordinating with the community and on the automated reporting (by the community to MAST) on the datasets used.

Suggestions for Future MUG Meetings

This year, the MUG had some suggestions to make future MUG meetings even more productive.

1. Make presentations available well in advance (a week is suggested). Advance presentations might shorten the actual time required talking about the slides and leave more time for questions and discussion. The advance time would also give the group a little time to contemplate possible issues. A little time spent in advanced preparation may make us (the MUG) more effective for you.

2. Presentations are often keyed to educating the users' group, which is good, but some of this education could be done in advance; presentations should clearly delineate where you might like input or suggestions from the users' group. If you anticipate a lot of time spent on a single issue or two, providing us links to suitable white papers or other background material might be warranted.

3. While a day is probably the minimum amount of time for a meeting, a day and a half or two might be better, to give some time (and thinking) between the presentations and the preparation of the report. It also would allow for more time on the executive debrief session. [Note added: The chair acknowledges that the single day constraint was a priority for her participation.]

4. Some time should be set aside for reviewing and presenting the previous annual MUG report, if only to update new members on the purpose of the MUG and the best ways to optimize the time spent by MAST and the committee members towards enhancing the scientific value of MAST and its services.

5. Most MUG members are not familiar with all aspects of the MAST, so it is important to provide some background about the roles of the various components of MAST (HLA, Missions, general archiving, etc), but a complete review of each component is not necessary. It should be possible to give a one-page summary of the role of the component, its level of effort by MAST. Then go straight to issues that were dealt with last year by the MUG, how things have or have not been addressed, and then proceed to current issues and plans. This recommendation is quite consistent with #4 above.