

MAST Users Group Report Dec 2014 MUG meeting

The MAST Users Group (MUG) met at the Space Telescope Science Institute on 2014 December 2-3. MUG continuing members are Guy Stringfellow (University of Colorado; Chair), Sarbani Basu (Yale University), Timothy Hamilton (Shawnee State University), and Letizia Stanghellini (NOAO). We welcomed new MUG members Cynthia Froning (University of Texas, Austin) and Todd Tripp (University of Massachusetts-Amherst).

Once again, this year's meeting format followed the revised format from last years MUG. The amount of presentations have been reduced significantly, focusing less on intricate workings of activities within MAST and instead presenting higher-level concerns for which the MUG's constructive feedback can be provided. The presentations were more interactive with the MUG, allowing questions and discussion to flow between MAST and the MUG. This format has proven quite successful and should extend to future MUG meetings. It is useful for the MUG to receive the meeting agenda presentations about 2 weeks in advance of the meeting so they can be reviewed and questions/concerns formulated for discussion at the meeting. These presentations do not need to be finalized in advance of the meeting, but should include the major points that will be presented and discussed. The final presentations should remain available on the MAST public 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 in these reports. Stringfellow collated and served as editor for this report.

At the end of the meeting the MUG elected Sarbani Basu as the new MUG Chair. After the close of this meeting and the submission of this report, Tim Hamilton and Guy Stringfellow completed their terms of service and rotated off the committee; new MUG members will be selected later, in advance of next years meeting. Tim and Guy thank their MUG co-members, the MAST staff, and all who presented and engaged in the many fruitful discussions over the years. It has been a pleasure serving and interacting with you all.

Overview Comments

MAST is currently at a crossroads, transitioning to a new MAST Portal website that will be introduced early next year that will change the manner in which data will be accessed. "On the fly" processing will cease and all data will be available as an online cache. With these impending changes will come improvements in performance in accessing data across all MAST mission holdings. We refrain on commenting herein on items that should be resolved or will no longer be relevant when these new changes are implemented, except where deemed important. Most of the MUG comments that follow are either in direct response to questions posed by the MAST Team, or that surfaced from the ensuing discussions during and following the presentations, and internal MUG deliberations.

The MUG would like to see a slide presented at the onset of the MUG meetings outlining the structure of the Archival Branch, highlighting those individuals in charge of the interactions

between the MAST and the community, and how the MAST personnel formally interact with other STScI branches.

Carryover from Previous MUG Reports

- Regarding the MUG 2012 and 2013 recommendations that the WFC3 persistence problem be tackled jointly by MAST (for archive data) and the WFC3 instrument team for application specifically to the MAST archive, we again strongly recommend this important issue continue to be developed. While a difficult problem, there are several reasons why this should be performed by STScI staff for the MAST public archive. The instrument and MAST teams collectively have the necessary expertise and knowledge of the nature of the persistence problem and how this should be applied to the WFC3 data. Just as important are the logistics concerning access to and ability to process the related data sets necessary to work out and apply the appropriate correction. Incorporation of a persistence correction into the standard pipeline is essential for the legacy value of WFC3 data. The 2014 MUG members encourage future members to continue to make this a MAST priority and see its implementation fulfilled. Providing both persistence uncorrected and corrected data, as a download option in the portal, will allow users to choose which data to work with and to evaluate the persistence problem impact on their science. The MUG is concerned about the ultimate quality of the data residing in MAST, and ideally archival data should have the highest quality calibrations and corrections in order to maximize the scientific impact of the archive and to avoid the often (human and computer intensive) time consuming repetition of the data reduction processing. We appreciate that MAST is engaged in the WFC3 persistence correction and taking an active role in this calibration issue, and we encourage MAST to extend these activities into other instrument/mission domains.

New MAST Archive Calibration Concerns - COS

The MUG would like to now raise a similar calibration issue that is very important for the long-term value of MAST data. This is the issue of the wavelength calibration of the Cosmic Origins Spectrograph (COS). It is well known that the COS wavelength calibration suffers from serious problems that can be greatly alleviated with some effort from STScI. COS has the same type of detector as the Far Ultraviolet Spectroscopic Explorer (FUSE), and FUSE had the same problems in its early days. The FUSE problems were rectified fairly quickly. The COS wavelength calibration issues have been discussed in several contexts such as the COS calibration workshop that occurred at STScI not long ago. The MUG considers this archival data quality issue to be a high priority for spectroscopic users of the HST archive, and we would like to strongly support the request that STScI invest resources in the improvement of the COS wavelength calibration. If this problem is not corrected, this will ultimately limit the scientific impact of this spectrograph. COS provides unique data, and it is unlikely that another space-based telescope will be able to provide similar data for many decades after HST is decommissioned. For this reason, the MUG considers this to be an important issue from the point of view of the ARCHIVE USERS because these data will be the only information about many important science questions for the next 30 years or more. The urgency of this calibration improvement has been elevated by the successful

ultraviolet initiatives in Cycles 21 and 22 - many new COS programs are now underway, and it is very important to improve this fundamental limitation.

Annual MAST Survey 2014 and Community Feedback

The 2014 MAST users survey was sent out via an expanded HST email list and all active archive accounts. The survey, and its importance, was advertised in the STScI Newsletter, on the MAST site, and various social media sites. Anton Koekemoer presented the survey results. There were 321 responses from a distribution of ~5,000, resulting in a 6.4% response rate. This more HST-user-focused targeting of the survey had a higher percentage response than previous years: up from 3.6% from the 2012 survey (301 responses from ~8,410 users), but only a marginal increase over the 2013 survey at 5.8% (180 responses from ~3079 users). Only 3% of the users (~9 people) responded as never using the MAST archive, while the same percentage responded as using the MAST archive daily. The vast majority of the respondents in 2014 (97%) indicated they have used the archive at least a few times a year. The MUG encourages continuing efforts to expand the response rate of the survey, and as in previous years we provide suggestions below. However, this response rate may remain relatively low and roughly flat in future years, and the question then becomes what is the value of the survey and is it providing the essential feedback to MAST needed to serve its community. The answer to the latter is quite apparent based on the survey results. Those responding are using the MAST archive! Perhaps they should be viewed as the core users of MAST, and hence the feedback they provide represents active experienced users utilizing MAST resources, and as such their input via the survey is important. It is likely there is a larger population of archive users out there, and it is also just as likely that many (most?) will never respond to the survey. Instead, the several hundred users who do respond to the survey will be the ones who will help transform the future functionality and utility of the archive.

There are many aspects of MAST that the survey seeks input on from the community. This is unavoidable. In previous MUG reports it has been noted that long surveys are perceived to be a tedious task by the community (even when they are not). This is compounded when surveys ask for lengthy narrative commentary. Narrative commentary should always be allowed, but not required to capture the essence of the survey question. Thus, the survey questions should be composed in such a way as to obtain a meaningful response by either a simple yes/no answer or via selection of a set of multiple-choice answers (with more than one answer allowed in many cases). Breaking the survey up into short segments and distributing these segments at different times throughout the year might improve the response rate.

Providing motivation to the community for completing the survey is essential to increase participation. Examples of the impact on important issues resulting from the community response to past surveys should be included in the STScI Newsletter.

We'd like to revive our recommendation from last year that the Portal incorporate floating windows with targeted survey questions when users use specific functions in the Portal. This might:

- * increase the survey response rate, because it's on people's mind when they're asked, and it makes it easy to do, and it's a *short* response.
- * sample a more representative set of responses, rather than the selection bias that the long survey today creates.

We recommended this last year, and the MAST response was skeptical, asking how often we bother to fill out exit surveys ourselves. True, but of the 6 MUG members this year, only 3 bothered to fill out the existing MAST survey. Surveys query researchers often at the wrong time – when they are not presently involved in the task being surveyed. This then requires a thought-shift that many are unwilling to (or simply can not) do at the moment. When immersed in a task that perhaps one is encountering difficulties or whimsically thinking how something could be done differently is the time to nudge the person and ask *can you put down that thought in writing*. Chances are the response rate will increase, and more substantial comments will be provided.

The Portal survey we're recommending could be done in the following fashion:

- * It would come up as a floating window, the same way the image preview windows are.
- * Its a single question, related to what the user is doing (if possible). Display the question directly in the window—don't give a link to it.
- * Multiple-choice questions could have buttons to select that requires making a selection for the user to continue, dismissing the window.
- * To minimize annoyance, the survey question would not come up every time the user takes an action. Allow them to pop-up randomly. For users accessing the archive with their own accounts possibly keep track of the user having responded to the question already. That means tagging the users account to keep track so they are not perpetually bothered with the same questions.

The 2013 MUG report may contain additional useful suggestions or nuances on the above.

New MAST Portal

The portal undoubtedly should become a very useful tool, and it was heartening to see that the portal is beginning to be used. We would like to reiterate some of the suggestions that we had made in the 2013 MUG report. These include highlighting through news articles announcing new datasets, data reprocessing, new ancillary products, and so on when they are available within the portal. The news tab is still drab and essentially invisible and hence the design could be revamped. We would, as another example, still like to see a feedback button to report problems on the fly instead of having to send an e-mail. In fact, if there is a way to capture the page and its recent history related to the problem being reported, this may alleviate some of the burden of reporting detail required by the user and assist MAST staff by providing relevant information that may elucidate the problems. Another issue is links to instruments and data processing manuals; last years report requested those to be included in the MAST, and MAST's response concurred. However, if the links are in the MAST they are not easily visible. Again, we would urge the MAST group to revisit the past few reports and the responses contained within them and ensure that what had been agreed on is actually implemented.

The MAST website:

The header has a tab called "Mission_search" which is extremely confusing. It has multiple options for the same mission without saying what difference there may be, if there is any. An example is "Hubble" and "HSTonline". At the very least, once one enters the search form there should be a statement describing where one is within the site, and what information is subsequently accessible being in this sub-folder. Other search options related to this sub-folder could also be summarized so that if this isn't the proper place to obtain the information, one knows where to go and can relocate there. The "Help" function actually does not help. We assume that HSTonline will be the search of the cached data, and once all the HST data are cached, one can remove the older version, but till that is done, this issue requires clarification. Other confusing labels are "Kepler Targets" and "Kepler Data". The MUG presumes that "Kepler Targets" is really the "Kepler Input Catalog," and if so it should be labeled as such. Perhaps these will all be redundant once the Portal is the only interface, but until that is the case, the mission search header should be more informative.

Hubble Legacy Archive (HLA) and the Hubble Source Catalogue (HSC)

The HLA is an invaluable resource for science, and the HSC is a long overdue asset of the HLA. We anticipate its debut with enthusiasm and excitement. Here we provide some feedback on early interaction with the products.

In the link between the HLA and the HSC:

- * When the HSC objects are circled in the interactive image viewer, clicking on a source brings up its entry in the HSC table. It would be good to be able to select more than one source (Option-clicking? Drawing a region in the image to select everything within the region?).
- * In the HSC table that comes up, there is no clickable link to any other information or resources on that object. It would be useful if clicking on an entry would bring up a list of images/datasets in which the source exists. We understand that the cutout image regions will be implemented in the future, which will be very good.
- * Another useful feature would be the ability to save selected HSC sources in a "notebook," like SDSS uses. This could be exported in CSV, txt, HTML, etc. formats. Saved object ID's could be then used later in an SQL search, for example.

For further development of the HLA interactive image viewer:

- * The line and column plots seem designed for spectra, rather than images. (E.g., there's a line fitting option.) A log/linear option for the flux axis would be helpful.
- * For images, some features like simple aperture photometry and contour plotting would be useful to give quick look data. This would be especially helpful for proposal planning.

Saturation is going to be a problem in some HSC applications. For example, in the ACS observations of globular cluster M13, the brightest stars are saturated in at least the F435W and F606W images, and there are no other M13 observations made in these filters.

The capability for users to reject saturated sources would be useful. Either through an automation switch they can turn on, or by some sort of flagging; this doesn't appear in the table. In some cases like this one there are no other (unsaturated) observations and this will limit what they can do with the data. Admittedly, there may be no way around this.

The Hubble Source Catalog's CASJobs interface could use science-focused search examples; right now the examples on the query page are mostly things like "publications by country." There are science examples in the Table description pages; some of these should be added to the query page.

The CASJobs Plotting function for the results brings up the VOTable plotting tool, but no data appear.

Data Linking/Data Mining/AstroTag:

Three talks concerning the areas of linking data to publications, developing more sophisticated data mining techniques in the future, and constructing an astronomy thesaurus for tagging observations and research projects were presented to the MUG.

The MUG supports the effort (led by MAST and the AAS) to provide DOI links in published papers to the data sets used in the research. There was some discussion of the level at which the DOIs should be applied (e.g., at the individual exposure level or larger data set compilations) but the general consensus was that linking the full MAST data set used for the work would add value to the content of the published science.

Similarly, the MUG is supportive of the effort to develop improved tagging of astronomical research areas. In particular, the commitment to adopting community-wide standards, for example within the framework of the Unified Astronomy Thesaurus, is laudable and should be continued. The AstroTag presentation focused on some of the challenges in developing a useful thesaurus, but it would be helpful if future presentations also addressed expectations of how this thesaurus will be used by the community, specifics of added research value, and examples of how this may be incorporated into the MAST system.

Finally, the MUG heard an open-ended discussion of future options for more sophisticated data mining algorithms. This type of data mining will become increasingly important in astronomical research in the era of very large data sets. For MAST in particular, the incorporation of the 2 PB PAN-STARRS1 archive and upcoming JWST archival data will increasingly rely on advanced search parameters, such as the machine learning and spectral and image template inputs presented. Because this effort is at an early stage, the MUG encourages MAST to engage in strategic planning early to focus the development effort for maximum return. This effort may benefit from the formation of a committee to solicit community input and develop a long-term plan. The MUG also encourages MAST to actively engage within the astronomical community (e.g., attending the LSST workshop in Tucson) and within the broader data science community, where development efforts of this nature are ongoing. The MAST effort can then be focused on front-end user interfaces

rather than replicating the back-end algorithms that are not astronomy-specific.

Kepler, K2, TESS, GALEX, PanSTARRS:

The MUG endorses the MAST activities directed towards these missions and their archival data.

The MUG would urge that MAST contact the KASC group periodically to determine if updates are available for incorporation within MAST. This serves to expedite these updates.

The MUG is pleased that MAST is already involved with the TESS team and are planning the archive before the mission is launched! As discussed in previous recent MUG reports, being actively involved during mission development will produce a better data archive and will see this implemented and served to the community quicker and more efficiently.

We are pleased to see gPhoton is being implemented.

The ingestion of PanSTARRS data into MAST is laudable. Handling the large data volume is a rehearsal for future ingestion of extensive JWST data.

Enhancing HST Spectral Products

The MUG appreciates that the 2014 meeting presentations included some new items as requested in last year's report. On the other hand, after last year's presentation by the Hubble's Spectroscopy Legacy WG, we were looking forward to hearing about their progress. The most important item planned by the WG was the combination of science-grade co-added spectra for the most used HST instruments and modes. The MUG is concerned about the progress of implementing spectral previews and other tools that facilitate spectral research for the MAST users. The preview tools offered for spectral analysis are still very limited. The MAST portal offers thumbnail images of 2D spectra only, which are not very useful to assess the utility of a specific spectral database to a science project. It would be really important that a simple display of co-added spectra could be accessed through both the MAST search page and the Portal.

We are looking forward to these implementations. We kindly request a complete presentation by the Hubble's Spectroscopy Legacy WG at next year's MUG meeting, including a progress report on the planned items, which in this year's presentations were described only in the introductory presentation, and very synthetically. We specifically would like to hear about plans to implement spectral previews and coarse analysis to be available through the different data access pages including the Portal.

There are other items planned by the Hubble's Spectroscopy Legacy WG, including future plans for HLSP items for the future of the HST Archive. The MUG would appreciate if the MAST group could refer to our prioritization and discussion included in the 2013 final report, under "Enhancing HST spectral products".