

# **MAST Users Group Final Report**

## **Submitted: 2014 January 17**

The MAST users group (MUG) met at the Space Telescope Science Institute on 2013 November 18-19. MUG members are Guy Stringfellow (University of Colorado; Chair), Sarbani Basu (Yale University), Ed Fitzpatrick (Villanova), Suvi Gezari (University of Maryland; in absentia), Timothy Hamilton (Shawnee State University), and Letizia Stanghellini (NOAO).

This year's meeting format followed the recommendations from last year's MUG. The number of presentations was reduced significantly, focusing less on intricate workings of activities within MAST and instead presenting higher-level concerns for which the MUG's constructive feedback could be provided. The presentations were more interactive with the MUG, allowing questions and discussion to flow between MAST and the MUG. The MAST team and the MUG felt this meeting format, including the two-day structure, allowed effective interaction and communication and should extend to future MUG meetings. While the presentations were made available ahead of the meeting, the MUG would appreciate receiving the annual presentations earlier, about 2 weeks prior to the actual meeting to allow enough time to review the material before the meeting. The presentations remain 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. Stringfellow collated and served as editor for this report.

After the close of this meeting, Ed Fitzpatrick and Suvi Gezari rotated off the committee; new MUG members will be selected in the Spring, well in advance for next year's meeting.

### **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 from 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 MUG 2012**

- Regarding the MUG 2012 recommendation that the WFC3 persistence problem be tackled jointly by MAST (for archive data) and the WFC3 instrument team, the MUG 2013 again strongly recommends this issue be addressed. While a difficult problem, STScI staff are best poised to deal with this more effectively than individuals in the community. Particularly as regards the proprietary aspects of the related data causing the persistence. Incorporation of a persistence correction into the standard pipeline may be premature (though we hope that significant progress can be made in 2014 to this goal), there should be enough data now residing in the archive to provide some attempt at correcting this in the archive data. Providing both persistence uncorrected and corrected data allows the user to evaluate and choose which data to work with. We concur that moving to the online-cache mode early in 2014 should allow significant progress to be made on this important task during 2014.

## **Annual MAST Survey 2013 and Community Feedback**

The 2013 MAST users survey was sent out via email to all HST proposers over the last 5 years and all active MAST archive accounts. Anton Koekemoer presented the survey results. There were 180 responses from a distribution of 3,079, resulting in a 5.8% response rate – up from 3.6% from the 2012 survey.

Surveys have limited capabilities and expectations in capturing the community's feelings and experiences in using MAST. The worry is whether the primary users of MAST even respond. As the survey is “anonymous”, there is no way of knowing given the response rate. The tremendous science value of and outstanding resources (data and related tools) contained within the MAST holdings are solidly recognized by the community. The concerns are how accessible are they (ease and rapidity) and how can they be improved (and with what effort) to yield even better science value. Informing the users in the announcement of the Survey how the survey results will influence MAST and feed back into better community service may entice more users to respond.

The MUG felt that a more user interactive approach would yield improved insight into these questions than simply an annual survey. By implementing various feedback tools embedded within the access portal, two types of user feedback could be captured during the process itself: problems encountered during the access and retrieval process, and useful feedback. The former provides immediate feedback for problems that could prevent data retrieval and use, while the latter could make the entire process flow more smoothly and provide the users with better data products. Allowing the user to provide such comments without leaving the portal access should result in more relevant information specific to the process to be provided at a time when these are directly encountered/realized – the concerns are fresh in the users mind.

Feedback tools can be implemented through popup windows at critical times during data retrieval/access that query the user to report problems and feedback, as well as at the end of the entire portal session. One or more “Help”-type buttons could also be located in and remain on each page of the portal session, so if an issue materializes along the way, the

user can click for instructions on how to proceed or report. Examples could be “Problem” and “Feedback” labels, which should be clear to most users. Depending on the nature of the problem, one or more of these features could redirect the user’s response to the normal helpdesk, and provide a priority flag. Such details in the implementation of the scheme can be worked out internally within MAST.

Understanding the value of particular data sets and the missions they are derived from is important, particularly for future mission planning. It would be useful to identify through access/download hits which data is being accessed most by multiple users. This differs from the approach of how much data is downloaded; some particular datasets downloaded in volume by only a few specific users (professionals and amateurs alike). Wide use by large numbers of users indicates a different type of science value that should be identified. In discussion about this point between MAST and the MUG, implementation should be possible within the new portal.

## **New MAST Portal**

Related to the use of MAST is how effectively its content is communicated to the community. There are two areas in which communication could be improved. One is making the new portal home page more attractive and informative. The first page should not contain entry information to search the archive, as most users would simply go directly to searching for their data and ignore other important information (currently blended into plain nondescript text and formatting). There should be a NEWS highlighted area that announces new datasets, reprocessing, ancillary products, and so on. Flashing colored text and/or boxes could highlight items of importance. This will force the user (at least the new user) to actually view and hopefully read introductory material of relevance to them and MAST. Providing useful information presented in an attractive manner (not just dense text) would hopefully stimulate the users to look for new information on a regular basis. Links to the instrument manuals, data processing manuals, and science papers describing ancillary data and/or tools should be provided at relevant places within the portal to complement the “one-stop shopping” mentality; everything at the users fingertips, just a click away. Finally, inform the community via email announcements to new data releases and other important information. Advertising in and coordinating the annual Survey and important MAST announcements with the STScI Newsletter should be done. Also utilize the AAS announcements, which also occur via email to the community.

The MUG was surprised to see the emailed announcement of the release of the new Portal the week after the meeting. We all were expecting release sometime in 2014. Given the extensive feedback and debugging that occurred during the meeting, and the suggestions appearing in this report, the MUG expected some further development and inclusion of many of these items before officially launching the new Portal, particularly having the classic archive features implemented. If these had been done during the short interim between the meeting and the announcement, the MUG would have appreciated being notified in advance. Providing a thoroughly test-bedded new Portal to the community provides a smooth transition.

### ***Additional comments:***

- Tutorials are very helpful. They should be presented as screen shots on a help page for the new portal and/or be video (webcast) presentations highlighting use of the Portal, its sub-domains and tools accessed by a link or button.
- All classic features of the current MAST data access pages need to be carried over and immediately implemented into the new MAST Portal. This includes features like viewing spectra available for download. Thereafter these features can be extended for the newer MAST holdings.
- All of the table columns and spectral features for all holdings need to be completely defined for the user via pull down information or linked pages describing these. Units need to be clearly listed. Computing of magnitudes and the magnitude systems used should be clarified.
- Pan-STARSS is a great addition to the MAST holdings. While it is understood that separate hardware exists for processing, access (downloading) of the data will occur through the present bandwidth available through MAST. If this includes images, it may over burden the MAST network and slow down data access for other MAST users. For such large time-domain datasets like Pan-STARSS, is it possible to secure separate fiber-cable links to access these data?
- All spectra should be viewable prior to download. GALEX and slit-less spectra need to be included in this fashion.
- All proposal target name formats should be searchable within the Portal. Further, discussion should be held between MAST and the Proposal development team to allow standardized (i.e., SIMBAD, NED) target name formats to be accommodated. MAST targets should be transparent and searchable under these names. As an example of the type of problems alluded to here are some of the planetary nebulae targets. HST proposal formats have not allowed in the past formal names such as PN Gxxx.x-xx.x. Proposers are forced to be creative, renaming their targets, such as PN-Gxxx.x-xx.x, which are not searchable under the beta-portal. All MAST targets should be cross-correlated with their scientific names, and perhaps having these listed in a pull-down or pop-up window is possible.
- Standardizing some features of the portal with those used daily on the internet is helpful (i.e., common terminology), perhaps changing the shopping basket to a shopping cart. Old items retained in the shopping cart should be differentiated from new items being added in the current session, using colors as an example.
- The archive should give descriptions of the formats of the different datasets. This is especially helpful with spectra, whose formats vary a lot from one instrument to another.

- The archive could give suggestions for basic software tools for users to view the data. An example is with GALEX spectra. They're in FITS files, but they can't be viewed with DS9 the way some imaging spectra can. Some users will be unfamiliar with the datasets and how to view them, and to use them users will need this type of information.

## **High Level Science Products (HLSP)**

HLSP are valuable resources to the community, as acknowledged through consideration by Time Allocation Committee (TAC) reviews and opportunities announced by facilities. Once formally approved by the TACs, MAST should liaison with the teams providing these and inform them of the possibility of having these data reside in the MAST holdings, and of the data format requirements. Working with the teams at the earliest stage of their programs will help ensure of a timely and smooth data ingestion process. Announcements should be made to the community (and statements incorporated into the teams' published papers) that these products are available through MAST.

### ***Additional comments:***

- Types of high-level products that could be of wide interest to the community are often produced through NASA and NSF funded programs. With NASA these are primarily associated with large observing programs obtained through the individual mission calls. Closer involvement with MAST during the selection process and performance periods to ensure these products reach the widest possible audience would benefit the community at large, and undoubtedly the teams preparing these datasets. Early decisions to include (or not) relevant data products and tools within MAST will help in disseminating the products much earlier and in ensuring that they are useable within the construct of MAST. As regards HST and JWST, MAST should be involved in the process at all stages and be in contact with the PI after selection.
- Liaise with ADS (and with AJ, ApJ, PASP, MNRAS, and A&A) for tables, spectra, and images of high-level science products related to MAST holdings. Standards need to be defined and assured in the data products.
- Liaise with ADS about putting direct links to HLSPs in their "On-Line Data" links. This is done with Vizier tables. As it is now, "On-Line Data" has a \*general\* link to MAST, followed by a link to the low-level data products in MAST. And when HLSPs exist, there should be a direct link to them.

### ***Responses to specific MAST High Level Science Product queries to the MUG:***

- *"Do the community think of MAST as their go-to source for HLSP, or do they just Google search and visit individual people's / team's pages? How can we change this mentality?"*

The community needs to be educated about the availability of these products. We have recommended ways above in which the community could share data products using MAST.

All other questions to the MUG in the HLSP presentation have been addressed by our comments and suggestions presented above in this section.

## **Kepler Repository and Issues**

Progress on past MUG comment *“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.”* We agree with the MAST group’s view that stitching together different quarters of data is still a research problem and there is no accepted way of doing so. However, python scripts already exist for stitching together the different quarters and we urge MAST to make these available with documentation. Nevertheless, the added science value of having a reliable means of stitching together different quarters of Kepler data is huge. While we understand that the MAST team should not be responsible for tackling this, and that the Kepler planet-finding team has no intention of pursuing this important endeavor, we encourage MAST to liaison with the Kepler Asteroseismic Science Consortium in providing these resources when available. In fact, high-level data products derived from this consortium effort should be ingested into MAST in a coherent and timely basis.

We note that the comment in the presentation that many community members prefer to extract their own light curves from the raw data is a basic concept that applies to any data set residing in the MAST holdings. In fact, if this is the case, then perhaps only the Kepler raw data should be archived – the MUG thinks not. It is acknowledged elsewhere (even in other MAST presentations) that the capability, and certainly the expertise, to reduce even HST data in the future is likely to be lost to most of the future community users of MAST. This is the idea behind creating a “best reduced” archive, ready for use by the community for many wide-ranging science problems. This is the proactive approach that should be taken with the Kepler data. As the primary Kepler mission has now ended, it will not be long before the cohesive Kepler team evaporates. Losing the strong asteroseismic consortium expertise as a team is also inevitable. This is the time for MAST to be proactive and liaison with these teams to insure that the scientific archive nature of the Kepler data is achieved, before these external resources disappear. This applies directly to producing the “best” light curves and stitching of multi-quarter data together – what the MUG perceives to be the longevity value of the Kepler archive.

The Kepler part of MAST is extremely useful as it is, but can be made even more useful to the community by including additional higher-level products. The MUG was pleased to hear that one of the innovations planned is to allow searches in the new portal with KIC ID numbers, and we urge this be implemented soon.

Further suggestions are:

- MAST should liaise with the Kepler Asteroseismic Science Consortium to ingest higher-level data such as “corrected” time series suitable for asteroseismic analyses and power-spectra.
- MAST should contact authors of Kepler-data related papers for higher-level products such as masses, radii, stellar ages, and such. These higher-level products are currently digested automatically by CDS and appear as a link in ADS. It would be useful if these data-products were also available on MAST, though this should not place a burden on MAST in doing so. Perhaps providing links within the data portal is the best approach here.
- A somewhat improved version of the Kepler Input Catalog will be published soon. MAST should contact Daniel Huber, the lead author of the paper, for the new catalogue.

***Responses to specific MAST Kepler queries to the MUG:***

- *“Enhance existing webtools and/or add new ones? Catalog and preview plotter (R. Thompson), integration of Pyke (M. Still)? NExSci precedence for offering analysis tools along with data (not just a data repository)”*

The MUG agrees and believes these tools enable enhanced capabilities by MAST users. This is in accordance with the MUGs view that MAST should be a self-contained resource providing access to these tools (even if just through direct Portal links).

- *“Improving data linkage and connectivity with Exoplanet Archive? “Two-body” problem where NExSci has all the transit data, MAST has the actual data and star information. Any improvement in linking the two is beneficial.”*

The MUG agrees that the two should be linked in some fashion, perhaps through links directly inserted to each website/portal.

- *“Lessons learned applied to K2 and TESS? Knowledge gained can and should be applied to K2 (if it happens) and TESS going forward. Development of tools and services for Kepler likely to be relevant to future missions.”*

The MUG concurs with this approach.

- *“What to add as Kepler transitions to a legacy mission? Which communities are underserved currently? How will people want to use Kepler data 5, 10, 20 years from now? Which data will they want (and will we have the tools for them to use it)?”*

It will be important to have the higher-level science products as Kepler transitions to a legacy mission. The time series will be in demand for the near future, and should be kept. See the MUGs previous comments regarding higher-level products throughout this report.

## **GALEX Repository**

Overall, the GALEX section of the new portal is quite nice. There are few outstanding issues regarding the GALEX data, except that the format of all downloaded MAST data needs to be clearly delineated so that the user knows how to preview and deal with the data.

- The capability to animate any region in the GALEX region to look for variation is a very nice feature. It would be good to extend this to all other imaging data for other MAST holdings.
- Since the original GALEX teams observing in the scan mode have not been successful in scientifically reducing and analyzing these data, it is the MUGs view that besides storing these data for public download, nothing else should be done with them inside MAST at this time. MAST has other more important immediate tasks to deal with.
- Provide a link to or insert within MAST a detailed description of the data reduction methods employed for the GALEX archived data.

## **Enhancing HST Spectral Products**

While there are some very good directions being taken regarding the HST spectral library, the MUG felt that current plans presented delve into the scientific analysis and interpretation of the data. The MUG believes that MAST should present the best data possible, deciding (with advisory panel and instrument teams endorsement) what is the best data to present. In general, this should be minimally processed to produce the overall best science use of the data. It includes techniques of combining spectra, removal of instrumental artifacts, and so on. However, when it comes to providing extracted products, such as equivalent widths and line flux measurements and spectral typing, this treads into the scientific “interpretation” domain and should be left to individual users to perform. If this is done as a science project outside the primary duties of MAST, then so be it; the MUG does not want to dissuade the production of useful science products. But MAST should focus on producing the best-reduced data to the community since much of the community lacks the expertise (and avenue of communication with others within STScI) to perform these fundamental tasks and decisions. Extracting multi-object spectra, combining spectra, providing target information as well as spectral tools are all tasks endorsed by the MUG for MAST to carry out. Referring to the presentation on this by Alessandra Aloisi, Recommendations 1 and 2 on page 4 of the presentation (the highest ranked priorities) are within the MAST domain and should be carried out. However, recommendations 3 and 4 fall more within the scientific interpretation and falls instead under an independent science project – that perhaps could be added later to MAST as a legacy project. Separating the two types of endeavors is important and needs to be transparent to the community as we rely on the expertise within MAST and STScI to make sound decisions regarding reduction techniques and their implementation to provide the best data products, not the scientific interpretation of this data.



- Having spectral tools that run behind the MAST portal, not having to download and install these on local machines, would be useful for the community.

***Postscript after the report was prepared:***

Stringfellow has seen the announcement of the new MAST Portal made in the AAS Newsletter and the excellent article section contained in the December 2013 STScI Newsletter. These are the types of proactive community outreach needed as discussed by the MUG during the meeting and the ideas presented in this report. However, the email “archive” texts that go out, which are simply text alerts, could be made more attractive in several ways – e.g., reformatting to a more pleasant format style and include an image representing the data being announced in the alert. For example, in December an alert was issued regarding availability of the UDF2 data. There was no description of what the science intent of the UDF2 was, where it was located, and no pretty picture of the phenomenal data.