

# **MAST Users Group Report**

## **Jan 2016 MUG meeting**

The MAST Users Group (MUG) met at the Space Telescope Science Institute on January 14 and 15, 2016. MUG's continuing members were Sarbani Basu (Yale University), Cynthia Froning (University of Texas, Austin), Letizia Stanghellini (NOAO) and Todd Tripp (University of Massachusetts at Amherst). We welcomed new MUG members Dan Foreman-Mackey (Sagan Fellow at University of Washington) and Mariska Kriek (UC Berkeley).

Meeting consisted of presentations by the MAST (and other) staff, followed by discussions. Although all presentations has been scheduled for the first day of the meeting, the MUG requested that more time be given to discussions of the presentations, and consequently, some presentations were moved to the second day. We thank MAST for accommodating this request. The expanded time for discussions with MAST was very useful and allowed us to understand the issues much better. The MUG appreciates that the introductory talks included charts with all relevant MAST people and sub-groups. The acronym list, though incomplete, was also much appreciated. Making the list complete will make it easier to follow the presentations.

At the submission of this report Sarbani Basu and Letizia Stanghellini will complete their terms of service and rotate off. Cynthia Froning will step down because of her commitments as a member of the STUC. Letizia, Cynthia and Sarbani thank their co-members, the MAST staff, and all those who presented and engaged in the many fruitful discussions over the years. It has been a pleasure serving and interacting with all of you. Three new members will be chosen before the next meeting of the MUG. The committee recommends that at least one new member be an expert in spectroscopy; another early-career member could also bring in added insight. A member of a ground-based observatory could also bring added insight. Todd Tripp was appointed the chair of the committee for the next year.

### **Overview:**

The MUG has a much more positive view of MAST and the efforts of its staff this year than it had in the last two. Clearly, the effort that is being put in is having a positive effect. The MUG was particularly pleased to see the development of the spectroscopic archive. We congratulate the MAST staff for their achievements.

The proposed Archive Mission Office is a welcome addition to the STScI managerial structure. MAST is the strongest (and best funded) archive in astronomy and thus serves a significant portion of archival data products in the field, and drives much of the R&D for data access and analysis so a coherent vision for STScI archive activities is important.

### **Topics not discussed:**

There were a few topics that were not discussed this year. The first was the survey. In previous years there was much discussion on how to improve the response to the survey. Although there was no formal discussion, MUG was of the opinion that the survey was too long; in fact one MUG member left the survey incomplete because of its length. The MUG

would urge MAST staff to look into the suggestions provided in previous years and use those to see if the response-rate improves, and present the result at the next MUG meeting.

The second topic that was not discussed was the question of AstroTag. We had heard a presentation about this at the 2014 meeting and were introduced to the challenges being faced. We did not hear anything formally about the progress of this project this year. We were told informally that good progress has been made. The MUG recommends that there be a formal presentation at the next meeting.

### **Our comments and recommendations on topics that were discussed:**

#### **The Portal:**

The MUG was extremely impressed by the MAST portal. The selection menu on the left and the preview of both spectra and images make it very easy to identify the correct datasets. The astroview on the right directly shows the field of view on the sky, and thus is of great value as well. The portal is accessible for everyone, and provides a great resource for professional astronomers looking for available data set, as well as for teachers using it for educational purposes. Even for the general public the portal provides easy access to Hubble's archive.

The MUG, however, believes that the following features could be improved: First, though there is a search box for the portal on the MAST homepage, there is no direct link to the page without entering the coordinates. The portal can only be selected using the pulldown menu under "Tools". Second, once added to the shopping cart, the actual download process could be made clearer. Instead of the "floppy disk" symbol in the pop-up window, the button could simply state "download" to avoid any confusion. Third, there is the issue of compatibility between postal file ingestion and SIMBAD output. The MAST portal would be more useful if it could read a variety of formats and tabs directly.

*API access to data:* There has been substantial work on developing the impressive MAST Portal as a common entry point for data search and access. This resource is valuable for many use cases but some users rely on programmatic access to the data. For example, it is useful to be able to directly query MAST from within an IPython notebook or another interactive programming environment and display views of the data that aren't natively supported by the Portal. Programmatic access can be obtained using the "MAST Web Services" product but this implementation is outdated, poorly documented, and the features it provides are limited compared to newer tools like the Portal. The MUG recommends the development of an Application Programming Interface (API) to expose the backend of the Portal for general use. This will allow the community to develop and support new tools (e.g. astroquery/astropy plugin) and use existing tools (e.g. kplr) for interfacing with the Archive. APIs are the industry standard for programmatic access to large online datasets and we expect that a MAST API would be broadly applicable and popular.

*Website redesign:* The MUG is enthusiastic about the proposed redesign of the website. One major benefit of this redesign would be an improvement of the "Google-ability" of pages on the site. As a provider of online resources, the usability of the MAST website is crucial and we believe that hiring a professional UI/UX designer to lead the redesign would be hugely beneficial.

### High-Level Science Products (HLSP):

MUG is pleased with the efforts being made to make the high-level science products accessible to the community. The amount of high-level science products collected on this page is impressive. The structure of the individual pages is however, not very uniform, and thus the available products vary greatly among the different project. The HLSP webpage could improve its organization, particularly by grouping by mission, by observing mode (imaging/spectroscopy/etc.) and by science topics. Additionally, several programs directly link to webpages outside the MAST. To this end the MUG discussed the utility of a HLSP template, to prompt users willing to submit their science results to organize and catalog their packages and findings. However, the MUG understands that receiving any type of high-level science product is preferred over getting nothing, and too many rules should not get in the way of scientist releasing their products.

As not everyone who uses the portal may be aware of all the high-level science products available, it would be helpful to have the systems be linked. For example the "observation title" in the left (selection) bar in the portal could link to the corresponding HLSP page, if one is available.

The MUG is concerned that HLSP appears to be under-supported (in FTEs) relative to the level of importance of the science output (as measured by data served and downloads).

### Spectral Legacy Archive

The news that MAST now has a (soon to be made public) Spectral Legacy Archive was received very well by the MUG. Clearly there is still work to do, in particular the listing by subject. We suggest that there should be a set of official astronomical fields/categories and keywords that are used uniformly through the STScI's system. For example, the keywords that can be selected for proposals should be compatible with those of the MAST spectroscopic archives.

We were surprised and somewhat concerned that many within the MAST team did not know too much about the development of this archive and that there are parallel efforts that mimic some of the functionality its functionality. The different teams should combine their efforts and minimize wasted time and effort.

### DOIs and versioning:

A prototype Digital Object Identifier (DOI) system designed to enable lightweight referencing of MAST resources with the goal of improved reproducibility was presented. The MUG recognizes the use and importance of this system and we are cautiously optimistic. Since the DOI concept is broad and flexible, there were some concerns about the semantics and DOI creation workflow. For example, should there be a single DOI referencing all the data used in a given paper or should the IDs be finer grain? Any issues like this are likely to be alleviated with clear documentation and a working implementation.

For the purposes of reproducibility, DOIs would be especially useful if they could also reference public datasets from other providers, in particular, ground-based surveys now that NOAO is working on the concept of a data lab with the goal of efficient exploration and analysis of large databases from 4m wide instruments. We believe that MAST has the

opportunity to demonstrate the benefits and encourage the broader community to participate. The MUG recommends further development of this prototype.

The discussion of DOIs brought up a related but deeper question about the versioning of MAST products. In principle, a DOI should point to a specific version of a dataset reduced using a specific software revision. In practice, this might not always be feasible—and the point was raised that the general scientific results in a paper shouldn't be sensitive to data reduction versions—but if a resource isn't versioned, this fact should be clearly documented and visible at the data search/portal level (i.e. not just in the FITS header).

### Big Data Science Definition Team Activities

The MAST team is looking forward to the hardware capabilities that will be required to support future data-intensive missions, particularly WFIRST. The big data presentation outlined the science use cases that were studied to develop the infrastructure requirements looking ahead, both for data storage and local data analysis. There have also been discussions with the Maryland Advanced Research Computing Center about the possibility of hosting the services there. This effort is endorsed by the MUG as a healthy long-term planning activity.

The upcoming Big Data Workshop is also a good, proactive opportunity to identify future data handling and analysis needs and educating the community on new methods of handling large data sets.

### JWST Science and Data Management System Overview

The run-up to JWST science operations is clearly occurring in earnest. The planning modes and Phase I materials for JWST will differ from previous missions and proposal template development is advancing. JWST is driving a lot of archive development —hardware and infrastructure improvements, new software tools, etc. — that can be leveraged by other current and future missions.

The MUG had no serious concerns about JWST development except for the short timeline for commissioning and between the end of commissioning and the Cycle 2 call for proposals. Although the FTEs may be in place to do the work on those timescales, the schedule is much more aggressive than has been pursued in the past (e.g., after HST servicing missions) and the institutional infrastructure will be taxed in a way it hasn't been before. The MUG recommends a thorough look at what will be needed not just in development but in the ability to approve and hand off materials between divisions within STScI to meet the development timeline.

The subscription service being developed for JWST would be a welcome addition to other missions as well to increase information exchange between the observers and the instrument support teams.

### WFPC3 persistence problem

The MUG was pleased to see that headway is being made with the problem of persistence in WFPC3 images. We followed with interest the presentation of the WFC3 new calibration packages that will be implemented in the near future, including WFC3 IR persistence model, improved flats and darks, and CTE correction. We recommend that both the PIs of the concerned programs and the general MAST users receive some sort of alert when these

correction will be available, and when the data sets have been processed with the use of new calibrations.

### Kepler, K2 and TESS

The MUG appreciates what is being done to archive the data from Kepler and K2, and hopes for a smooth transition to TESS. MUG has no reservations about what is being done.

### PanSTARRS

The MAST effort to host the Pan-STARRS PS1 archive is nearing completion. The archive will be a useful community database and leverages the MAST expertise in hosting and serving data to support a powerful dataset. The MUG was pleased with the level of development and user access support for the Pan-STARRS survey.

The current plan on waiting a short time for the Pan-STARRS team to deliver the final catalog before releasing the image data is a good one, but if further delays occur, the MUG recommends going ahead with the release for the benefit of the community.

### gPhoton

The MUG welcomes the development of the gPhoton database of GALEX time-tag events. The ability to create light curves and images screened by region, time, etc. will enhance the science return of the GALEX archive. The current Python-based query code is simple and powerful. The MUG would be interested in seeing gPhoton expanded to include a simple web-based access mode for scientists who are unfamiliar with Python programming.

### Cosmic Origins Spectrograph Wavelength Calibration

The work carried out by the COS team over the past year to improve the wavelength calibration of far-ultraviolet spectra was summarized for the MUG. The presentation showed how the data-products obtained with the G130M and G160M gratings were improved and the efforts involved (1) empirical assessment of the errors in COS spectra by comparison of COS observations to observations of the same object with the Space Telescope Imaging Spectrograph, (2) characterization and development of corrections for detector “walk” in both the x and y directions on the detector, and (3) study of the role of geometric distortion and improvement of the geometric distortion correction.

The MUG was very pleased to learn that this problem is now the top priority of the COS team and to hear about progress in this effort. Detector walk clearly causes significant errors in the wavelength calibration and geometric distortion may be less important, but we applaud the COS team for evaluating both problems. However, some data presented in this summary show substantial and discontinuous offsets between the wavelengths of spectral features recorded with both COS and STIS, and it is not clear if the origin of these problematic regions is fully understood. A better detector walk correction may fix some of these problems, but ultimately it could be necessary to develop an empirical correction to the wavelength calibration as was done with the Far Ultraviolet Spectroscopic Explorer, which had very similar detectors. The MUG urges the COS team and STScI to continue to vigorously attack this important problem, and we would like to hear another report on progress on this effort at the next MUG meeting.

## Feedback to the STScI Director and STScI Priorities for the HST Archive from the MUG Perspective

As the Hubble Space Telescope transitions into its twilight years, it is more important than ever to invest resources in the calibration of the HST data in the Mikulski Archive. After HST is decommissioned, the telescope will continue to have a major impact for many years because the archive is such a tremendous source of rich information. However, to maximize the impact of the HST archive, the data must be very well calibrated (hopefully it will be optimally calibrated). The HST archive includes a bewildering variety of data from many multimode instruments, and proper calibration of the archival data is a daunting task. At the present time, STScI has adequate resources (and, more importantly, the necessary high-level expertise) to work on this crucial effort, but after HST is decommissioned, it is certain that these resources and expertise will dissipate, probably rapidly. On this basis, the MUG believes that it is very important to make optimal calibration efforts a top priority at STScI at the present time.

The MUG discussed the calibration needs of HST instruments with STScI Director Ken Sembach. The director asked us to express our priorities in this report, so we state here that optimal data calibration for the final HST archive is absolutely crucial. This is particularly important for the ultraviolet data, especially the spectroscopic ultraviolet data. At the present time, NASA and space agencies of other countries have no formal plans to deploy an ultraviolet successor to HST. Consequently, the ultraviolet data from HST will be the most useful/powerful data for many purposes for at least 20 years. It is possible that some types of ultraviolet data will be obtained with explorer-class missions that can be deployed more rapidly, but high-resolution spectroscopy requires a telescope with significant collecting area, so the spectroscopic ultraviolet data will remain unique for a long time and will not be supplanted by new data from lower-cost missions. Given the uniqueness of these data, it is important to ensure that all of the UV observations are adequately calibrated. Ideally, this calibration should be the best achievable calibration, but at a minimum the calibration should archive the specs that were initially promised for each mode of the UV instruments. Currently, there are examples of UV spectroscopic data that are not even calibrated well enough to satisfy the original proposal specs. This is partly due to lighter demand for these modes. However, even though current demand for the near-UV mode of COS, for example, might be less than demand for other capabilities, the near-UV COS data will be the only data of its type for many years, and the value of these data are likely to grow considerably when HST users turn away from guest observer programs and focus their efforts on squeezing the most out of the archival data.