

MAST Users Group Report

March 2019

The MAST Users Group meeting was held February 27 – March 1, 2019 at STScI. The members of the group in attendance were Knut Olsen (Chair; NOAO), Tabettha Boyajian (LSU), Martha Boyer (STScI), Ben Montet (U. Chicago), and Karin Sandstrom (U. California San Diego). John O'Meara (Keck Observatory) was unable to attend.

This report summarizes the MUG's impressions from the presentations given by MAST staff and the discussions that they prompted, and includes some specific recommendations throughout. Whenever possible, it addresses questions posed directly to the MUG through the presentations. While not all questions could be answered, the MUG appreciated having these questions to focus their attention.

The MUG was very impressed by the significant accomplishments, across a wide range of areas, that MAST has made over the past year. MAST provides a critical service to STScI and the community, and is doing its job very well. MAST has undergone significant restructuring throughout, with the result that its goals and vision, its presentation to its users and the outside world, and its core services and teams all appear to be very well aligned. MAST has demonstrated ability to handle very large datasets and designed a cohesive set of interfaces for many different levels of users, including giving programmatic access to its APIs and developing server-side analytical capabilities through a Jupyter interface. While its restructuring is not yet entirely complete, MAST appears well positioned to handle the challenges of the coming decade.

The MUG noted that MAST's ambitions for development likely exceed available resources, such that prioritization will be needed. The MUG hopes that its recommendations and impressions in this report are helpful to MAST in making these priorities, or in requesting additional resources to meet its needs. One area that the MUG identified could use additional targeted resources is serving the needs of the Solar System science community.

MUG members Martha Boyer and Knut Olsen are rotating off the committee. The MUG proposed that Tabettha Boyajian and John O'Meara serve as co-chairs for the next MUG meeting.

Introduction

The MAST User's Group (MUG) Meeting opened with a review of the structure and purview of MAST and a presentation and explanation of the MAST Mission Statement, which is to "Maximize the Scientific Accessibility and Productivity of Scientific Data". As demonstrated in the presentation, more than half of the publications that derive from STScI-related data are from

archival use of that data, which makes MAST the most productive part of STScI's mission. The MAST PI also presented the results of a bibliographic survey that demonstrated that the rate of archival publications compared to GO publications is significantly higher at smaller institutions with lower numbers of total STScI-based publications, and even more so in countries with lower GDP per capita. The bibliographic survey also showed that the growth over time in GO publications at smaller institutions has begun to mimic that at larger institutions, but with a delay in time compared to the growth in archival publications. The MUG interpreted this behavior as showing that archival data usage allows underprivileged institutions to get started using STScI data products, and later become successful GO applicants. MAST thus serves an extremely important role for STScI, as it both maximizes the productivity of its facilities and makes those facilities increasingly accessible to a broad segment of the community, as captured well by the MAST Mission Statement. STScI should consider this dual impact of MAST when evaluating its overall priorities for future investment.

The MUG recommends that MAST continue to pursue its efforts to boost accessibility of STScI data. Specific suggestions include:

- Conduct in-person visits of MAST staff to smaller institutions across the nation to gauge their data archive needs
- Have representation from smaller institutions on the MUG
- Initiate discussions with other archive centers about their approach to measuring and increasing accessibility, and to coordinate accessibility efforts

PanSTARRS1 DR2 & Catalogs

The MAST Team presented a summary of the PanSTARRS1 (PS1) second data release (DR2), which includes the single epoch images and photometry of 3π steradians of sky, containing 11 billion objects (73 billion detections) and yielding a 1.5 PB image collection and 150 TB database. While not funded by NASA, the project has allowed MAST to develop the capability to comfortably serve extremely large datasets and address the scalability and performance of its design decisions. The PS1 DR2 release included a new streamlined catalog search interface, which will serve as a prototype for all future catalog search interfaces, a new TAP interface to the catalog, and the distribution of Jupyter notebooks with examples of scripted API-based access to the images and catalog data.

The MUG congratulates the MAST team on this highly significant accomplishment, and strongly endorses the approach encapsulated by the new interfaces, in particular the new catalog search tool and the Jupyter notebook examples. The MUG recommends that MAST begin gathering data on the usage of the different services (e.g. TAP vs. CasJobs), to gauge their impact and utility. For additional PS1 data products, the MUG suggests that MAST consider the Medium-Deep survey and the Pandromeda survey as having highest priority.

TESS & Exo.MAST

The MUG was presented with an update on the development of Exo.MAST and the rollout of TESS data to the community. The MUG was very impressed with how MAST handled the initial public releases of TESS data, as well as the community support they have provided, both online and through workshops, and feel that the team should be commended for their efforts.

The MUG felt that, with the many ways to access TESS data across MAST, the team should consider strategies (additional notebooks, tutorials, etc) to reduce the initial barrier to entry for as many as possible. Obtaining additional statistics on which access methods are most used, where possible, was seen as a benefit to help MAST allocate resources in the future and is encouraged. The use of AWS was seen as a benefit to make the large datasets more accessible, and the development of a visible “AWS 101” tutorial for astronomers was seen as a need, as well as additional tutorials, perhaps broken down by science case, that also present ways to leverage tools like AWS in significant and useful ways.

The Exo.MAST rollout was seen as very successful. The search interface, in particular, was seen as very clean and easy to use. The MUG anticipates that additional flexibility in tools (for example, the possibility to fold on an arbitrary period and phase in the phase calculator tools) would provide a significant benefit to the community. Incorporating links and building synergies with other tools (exoCTK, community TOI lists, and other widely used community tools) would enhance the perceived value of Exo.MAST beyond a tool for JWST proposal preparation.

HLA (Hubble Legacy Archive), Hubble Source Catalog, and Migration

The improvements to both the astrometry and the photometry of the Hubble Source Catalog (HSC) are a great achievement, providing a high-value resource to the community. The MUG was curious how much the HSC is used by researchers and encourages MAST to explore whether its visibility can be increased via additional community outreach. The example of the Bulge work from Calamida et al. is impressive and the HSC would benefit from providing these sorts of examples to the community.

NAVO Progress & MAST Strategy

The MAST Team presented an introduction to the VO ecosystem and the NAVO project, and reported on its progress and strategy on NAVO initiative. The presentation aptly described the VO effort as historically having operated on an island separate from core MAST work, which MAST has made significant efforts to mitigate. In particular, the MUG recognizes the excellent work that MAST has done in integrating select VO protocols into its services and exposing them through AstroPy-based APIs. The MUG recommends that MAST continue to engage in efforts

to bridge the traditional VO and Python in Astronomy communities, as there is significant development effort to be leveraged in doing so.

MAST Outreach

The MUG appreciated all of the effort put towards increasing MAST outreach, particularly in hack sessions at conferences/workshops, in promotion through MAST evangelizers, including news updates on the Portal, advertising on social media, and in policy changes to archival proposals. We encourage MAST to continue these efforts, with emphasis on continued diversity in outreach activities to reach the broadest possible cross section of the community. The MUG also strongly commends the effort put into the MAST Newsletter, which has a clear impact on community knowledge, as demonstrated through the timing of PS1 DR2 downloads, for example.

The website redesign is much improved over the previous incarnation, and provides an additional tool for educating the community on the available MAST resources and tools. However, the MUG thought the website could use a little more thought on its organization. For example, the PanSTARRS catalog query tool is not easy to find, and navigating to the Portal is not entirely straightforward. MAST might benefit from applying its first tenet of MAST interfaces to the design of the website (Know the Mission).

Tenets of MAST Interfaces

The presentation to the MUG included a discussion of the design tenets of the MAST user interface. Given the ambitious development efforts that have been occurring in various areas of MAST, the MUG felt that laying out these principles was a useful activity. The six tenets are:

- know the mission
- risk is the soul of innovation
- leverage shared experience
- anything added subtracts
- friction is the enemy of discovery
- polish to the mission

There was an interesting discussion of the fact that several of these tenets can be viewed as being in tension with each other, and how that tension was created on purpose (e.g. “risk is the soul of innovation” versus “anything added subtracts”). The MUG noted that these tenets will be useful in clarifying some of the development goals that were presented at the meeting. For instance, the MUG felt that “Know the Mission” is an important tenet to apply to the list curation framework discussed at the end of the meeting.

Notebook Strategy and JWST LRD work

An overview of the Jupyter notebook interface to MAST was given to the MUG. The MUG recognized that notebooks are appealing because they are executable, enable reproducible analyses, and are an excellent platform for combining code, prose, and visualization. They support over 40 languages (with Python being particularly important for the astronomy community), are open source, and are used outside of the field of astronomy. These aspects of notebooks all follow MAST's mission statement to enable discovery through many levels of expertise in the community. The MUG was then given a more specific summary of how notebooks fit into MAST mission priorities for both current and future projects (e.g. WFIRST and JWST). These include enabling community science work, collaborations, user support, pipeline development and distribution to the community, as well as an excellent way for more complex data analysis procedures to be captured (e.g. tying them to HLSPs). The MUG was presented with a strategy on how the "infrastructure" of notebook collections would work. The strategy included a repository of notebooks to which users can contribute, a style guide for those contributions, a set of Docker container images to support needs of users on different platforms, and automatic builds to ensure functioning notebooks. The MUG recognized that the ability to work in Jupyter notebooks and run software through Docker containers in tutorials has been well received by the community, for example in the TESS workshop a week prior to the MUG meeting. The MUG thought the implementation of live notebooks was an impressive tool for research discovery as well as to bridge the gap in community expertise using notebooks for their research.

The MUG is extremely supportive of the development of Jupyter notebooks as an analysis platform and as a way to expose MAST APIs. The MUG also notes that not all of MAST's users are comfortable with the notebook environment or with Python, so that MAST should take steps to make sure that those users don't feel left behind as the notebook environment is further developed. The MUG recognizes that remaining action items to be addressed include the specifics of the workflow for contributing to notebook development and making it easy for users to follow and use such tools. Along these lines, MAST should continue to consider how to maintain the curation of notebooks and users' compliance with the style guides, and how to help MAST users find the notebooks that they need. Suggestions for doing this include notebook galleries (<https://observablehq.com> is a nice example), a searchable widget through keyword search, or sorting by popularity.

Toward a Unified MAST Search/Future of MAST Classic

The MUG was presented with a future model of MAST search, incorporating a "one-box" model from which each of the traditional forms could be searched. The MUG was impressed with the

integration of the different access methods into a single interface. The MUG particularly appreciated that changes through the web form or search bar simultaneously changed the other, and the synergy between the search bar and astroquery API for ease of learning both tools. The MUG also was enthusiastic about the direct button to create a Jupyter notebook showing the identical search through astroquery. The group feels that this plan is an extremely promising model for the future of MAST search that should be pursued. There were concerns about the potential disruption in user experience in the future roll-out of these tools, and encourage the MAST team to consider the roll-out of the modern ADS search in their efforts to ease the transition for the broader community. The MUG felt that the timeline of the development of this search form, the integration of select Portal features, and most significantly, the timeline for the eventual migration and discontinuation of MAST Classic need to be defined as soon as possible.

MAST under the hood

As was apparent in many of the presentations given to the MUG, the MAST Team and its product have undergone significant reinvention over the past year – to great effect – as seen in the rollout of Exo.MAST. The presentation on “MAST Under the Hood” gave a glimpse into these changes, both in system architecture and in code development practices, at a technical level. These changes have included 1) the move to an architecture based on microservices, which has improved the performance and scalability of MAST services, and allowed for easy reuse and remixing in the design of new interfaces, 2) the adoption of a Continuous Integration/Continuous Deployment framework, which has allowed for rapid, robust growth, and 3) the adoption of an Agile development model, which has linked the change in code architecture and design to a set of development practices that readily accommodates the rapid development of modular code.

The MUG strongly endorses this reinvention strategy, and recognizes them as fundamentally linked to the new look and feel of MAST. The MUG also acknowledges the challenges voiced by the MAST team in making these significant changes to the code architecture and team culture. A concern raised by the MAST team itself was a feeling of losing sight of the strategic vision. Although viewed from the outside, it appears as if the MAST mission and technical development work are very well aligned, the MUG recommends that MAST hold periodic full team meetings to reaffirm and discuss the strategic vision. The MUG meeting itself in part has served this purpose over the past year, but if made the only venue for strategic discussions, the meeting co-mingles, and risks confusing, the dual needs of getting internal team buy-in on the direction of MAST and getting a review of MAST by external community members.

The MUG also notes that for many MUG members, the technical concepts and jargon made the “Under the hood” presentation difficult to follow. The MUG suggests that the MUG makeup should always have diversity along the axis of technological familiarity, and that detailed

technical descriptions should be accompanied by as much background and lack of jargon as possible.

The Extragalactic Face of MAST

Building on the efforts to create Exo.MAST, there was a discussion of future efforts towards an “xgal” or “highz” MAST which would unite HLSP, catalogs, multi-wavelength cutout images and more for the extragalactic deep field datasets. These products are currently heterogeneous and there are a variety of web interfaces designed by specific project teams to interact with the data. This means that there is limited discoverability across datasets. The effort has a well-defined mission statement: “To enhance access, increase discoverability, and enable visualization of datasets related to galaxy evolution from the MAST missions and aid in HST and JWST observation planning.” The team has done staff interviews and a “trace study” to inform their planning for this tool.

The MUG was strongly in favor of this effort. The MUG thought that the idea to unify the interface with the new search bar and the Exo.MAST was a very good plan. The MUG felt it would be worth expanding the user survey to other institutions (in the interest of scientific and research diversity). The idea of a “Common Catalog Archive Model” seemed like an interesting avenue to pursue in dealing with heterogeneous catalogs. The MUG thought that the goals and the scope of the project seemed appropriate. It would be worthwhile to see if links with other multi-wavelength holdings for these fields could be made (e.g. Herschel, ALMA, etc. observations).

The Solar System Face of MAST

Building on the subsequent discussion of whether there should be a “Solar System.MAST” along the same lines as Exo.MAST or Highz.MAST, the MUG felt it was important to emphasize that a substantial fraction of HST proposals focus on Solar System science, and that MUG members have heard feedback that this community feels somewhat ignored by MAST. Implementing an ephemeris tool and cleaning metadata would go a long way towards a useful interface for solar system objects. The MUG thought it was an important goal for the MAST team to investigate how to support the Solar System community and how to get the needed expertise into the team (e.g. by a targeted STScI hire or a working group, etc).

HLSP Update

High Level Science Products (HLSPs) are community-contributed data products that complement or supplement a MAST-supported mission. They can be custom-created products from those missions, or products from other telescopes used to complement MAST mission data. MAST has HLSP guidelines for users and contributors. MAST also provides a permanent

home for the data, permanent links to it (via DOI), as well as increased visibility through announcements and search interfaces for discoverability. Use of HLSPs is competitive to the other MAST flagship mission-produced data products. In 2018, eighteen new HLSPs were ingested in MAST, and 12 were updated (comparable to 2017 with 14 and 13, respectively) across a diverse set of data types and science disciplines.

The large initiative moving forward is the migration to Jahia Webserver, and the new HLSP homepages. The MUG was impressed with the new page organization and features, and we believe Jahia is a good choice with respect to the workload in ingesting the deliveries of new HLSPs. The MUG thinks the use of ELK for download/request statistics is useful and will help understand how the community accesses and uses HLSP products. The MUG thinks the development of version control, software to ingest HLSP to Common Archive Observation Model (CAOM), as well as new user search tools though catalogs are important, especially as missions (and their products) grow. The MUG thinks that the initiative to archive synthetic data is a useful and ambitious effort, and suggests that MAST consider how to integrate them with the aforementioned search tools.

On the question of the scalability of the platforms and search tools, the MUG did not have a strong recommendation to make. The current table of short descriptions with links to the HLSP webpages and the ability to filter HLSPs according to selected keywords appears to be adequate for the foreseeable future, but the MUG recommends keeping an eye on the usage statistics and soliciting feedback from the community to ensure that users are discovering the HLSPs.

JWST-MAST Spectroscopic Initiative

The information-gathering effort for the JWST-MAST spectroscopic initiative is impressive, including the consciousness of lessons learned from the different activities (particularly the AAS poster). The MUG recommends continued use of this approach for this and other efforts.

This initiative is in the early stages, and the MUG encourages MAST to continue considering the scope of this project, which has the potential to grow large and/or complex. The MUG recommends developing a plan and/or a demonstration that would need approval from the various stakeholders before work proceeds too far.

At present, this initiative appears to be heading towards the development of an additional user interface and the MUG believes this could be a useful tool for the community if given the same amount of thought as other new interfaces (e.g., Exo.MAST). Integration with existing STScI spectral data analysis tools (in whatever form they ultimately take) is essential for avoiding duplication of effort and/or community confusion. A consistent UX between what can be done in the Cloud and locally would be beneficial. The MUG felt that MAST should put some thought

into how many separate interfaces they are capable of maintaining, while noting that the decision to transition to an architecture based on microservices will make the operation of multiple interfaces easier to maintain.

Hosting Curated Lists

The MAST team presented the development of a repository of lists as a possible future component for MAST to support. Although the proposed solution has merit, the MUG was concerned with several aspects of this idea. Primarily, an open-source list of user-submitted and curated lists was perceived as being contradictory with the “brand” of MAST’s work hosting official data products. Where HLSPs require an accepted paper before being hosted, under this model any person could upload any information onto a MAST-supported page with minimal oversight, which was seen as a risk. Additionally, the MUG had questions as to whether these tools and the support of this framework, which may include lists of objects from facilities other than those supported by MAST, should be in the purview of MAST. The MUG felt that any future work on this project should only be undertaken after the careful development of a comprehensive plan and mission statement that acknowledges the scope and aims of the list project. The MUG also encourages MAST to discuss the project – and potentially collaborate – with other list curators, such as VizieR, in the development of this plan, as well as the potential user base.

Your Moment of Zen

One MUG member offered this poem constructed by direct quote from the “MAST Under the Hood” technical presentation:

*Restful, stateless, streamlined, agile.
Broader in our concept of identity, more fluid.
Federations
 fall by the wayside.
We don't really need the extra bits.*