Report of the 2020 Mast Users Group (MUG)

T. Boyajian (LSU, Co-Chair), J. M. O'Meara (Keck, Co-Chair), B. Montet (U. NSW), B. Nord (Fermilab, Kavli), K. Sandstrom(UCSD), M. E. Schwamb (QUB)

HOW IT STARTED

```
In [1]: import astropy.units as u
        import glob
        import numpy as np
        import matplotlib.pyplot as plt
        import os

        from astropy.io import fits
        from astropy.table import Table
        from astropy.units import Quantity
        from astropy.coordinates import SkyCoord
        from astroquery.gaia import Gaia
        from astroquery.mast import Observations
        from astroquery.sdss import SDSS

        from ccdproc import ImageFileCollection
        from IPython.display import Image

        from drizzlepac import tweakreg
        from drizzlepac import astrodrizzle
```

HOW IT'S GOING

This report summarizes the findings and recommendations of the MUG following the October 26th and 28th meeting, held virtually. The MUG were briefed by STScI/MAST staff, and an initial set of questions were submitted to and answered by J. Peek during the course of the meeting. This report will present the MUG's impressions of the materials presented, the overall trajectory of MAST, and its relationship to the MAST community.

The state of MAST

The state of MAST is strong. As an ecosystem, MAST delivers on its mission to maximize the scientific accessibility and productivity of scientific data by providing not only simple access to data, but a wide suite of tools for data discoverability, and increasingly, data visualization and analysis. The proof for this validation is evident in the growing rates of publication of archival data from MAST, which now usurp those from GO use of new data. Increasingly, MAST is also ingesting and serving data from missions and programs beyond those 'housed' at STScI. Two key examples are the TESS and Pan-STARRS repositories which are receiving considerable community attention and use. Likewise, MAST recognizes that accessibility is a critical component for engaging a wider, more diverse audience of users.

It is important to note that MAST's meeting its mission has been maintained during the very challenging environment created by a global pandemic.

Finding: MAST has risen to the significant challenge presented to the organization by COVID-19 and associated issues, such as work from home. The MUG applauds the MAST team for their ability to provide consistent quality service to the community despite these challenges, and for continuing to meet their mission.

MAST is not only developing for missions it serves, but is also cognizant of the future data horizon that will eclipse the current archive in both size and complexity as missions such as JWST and *Roman* begin operations. By planning now, MAST is well positioned to meet its mission in the coming decade. MAST is also beginning to meet the challenges of mission closeout, with large and complex missions, namely *Hubble*, on the horizon. With the future on the horizon, and the past looming, MAST's largest challenge is likely not to be innovative thinking, but the resources required to translate that innovation into execution.

Senior & Programmatic Review

MAST, by virtue of funding through the Astrophysics Data Curation and Archival Research (ADCAR) line at NASA, is subject to programmatic review every ~five years. In addition to review materials, inguide and overguide proposals are made for continuing funding. The MUG was presented with an overview of the topics presented to the review panel, with a natural grouping between past accomplishments and future plans. Many of the topics in the future plans section were covered in detail in later presentations to the MUG throughout this meeting. In summary, the result of the programmatic review was that MAST was excellent and meeting its mission.

Finding: The MUG congratulates MAST on a very successful programmatic review.

The review identified only one major area of concern, focused on cloud computing. The MUG has expressed concern around the cloud strategy in the past, and shares some of the same concerns expressed in the review. The MAST response to the review regarding current cloud use is sufficient in the near term, but a better description of the MAST cloud strategy as part of the larger STScI strategy would help further address concerns.

Finding: MAST currently does not depend on the cloud to execute its mission, but future missions and capabilities may require increasing dependence on the cloud.

Recommendation: MAST, in conjunction with STScI, should summarize the STScI cloud strategy to the MUG.

The programmatic review also highlighted a number of minor weaknesses. Nearly all these weaknesses are the focus of active work at MAST, and some of that work was described in later presentations throughout the MUG meeting. Specific MUG input will be given below on some of these topics. The MUG expresses confidence that concerns are being considered and addressed appropriately.

Finding: Areas of minor concern identified in the programmatic review are receiving attention and are on a path to improvement.

The sections of the programmatic review relating to future endeavors largely overlap with presentations given to the MUG throughout the meeting, and will be addressed below.

Computing

The MUG was presented with the status of the development of tutorial Jupyter Notebooks developed by MAST or end users and to be hosted at an official Space Telescope GitHub repository. The emphasis on improved rendering of these notebooks on the web was seen as a strength to overcome the perceived barrier to entry of viewing these tutorials through a GitHub URL.

Finding: Jupyter notebooks are well-used in astronomy, notably in the processing of HST and Kepler/TESS data, and their development has been seen as a strength in lowering the accessibility barrier to these data.

Recommendation: The MAST team should continue to explore opportunities to make notebook resources as accessible as possible to the broader community, not just those who are already familiar with notebooks and GitHub.

The ability to download these notebooks and use them as a pseudo training manual for notebook-based development is also seen as a benefit. The MUG raised issues around reproducibility in these resources, especially as they concern versioning.

Recommendation: MAST should develop a clear style guide for the contribution of notebooks to the MAST ecosystem that ensures these packages can be run remotely. This should at minimum include explicit calls of versions of each imported package and key dependency, although could be more extensive, up to the creation of a virtual environment or set of environments designed explicitly for each notebook to be run in the cloud, analogous to the T.I.K.E.

There was some discussion about how to create more tutorials and the value of outsourcing to volunteers, such as users supplying HLSPs, or hiring students and postdocs to develop tutorials at software development market rates, similar to the approach employed by the lightkurve team.

Recommendation: MAST should, possibly with solicited suggestions from the community, identify a 'complete set' of tutorial notebooks and project the total work effort required to build and maintain these notebooks in order to determine the optimal path forward towards constructing these notebooks.

For these tools to remain reproducible for the foreseeable future, continuous integration is essential. The MAST team briefly discussed their efforts to ensure these notebooks are a part of a CI scheme; this should be commended and should not be lost in future changes to the MAST ecosystem.

The MAST team also presented a demonstration of the Specviz tools under development. The MUG viewed these tools favorably as a step forward towards the development of cloud-based tools for the quick-look analysis of e.g. JWST data, and look forward to the future deployment of these tools.

The MAST team also described a series of current and planned science platforms to further their mission. They presented lessons learned from the STScI prototype science platform, namely that simply hosting computing resources and the data in one location (through AWS or other similar tools) was insufficient to meet the needs of the community. The use of the TESS Data Workshop Platform in 2019 was seen as a success for that conference, and future platforms are being considered for future development up to the Roman Science Platform in the mid-2020s.

Finding: The existing and planned platforms are successfully enabling the furthering of the MAST mission. These platforms should continue to be developed and made available to the community. Where they will be essential for future missions (e.g. exploration of Roman data), these platforms should be made available to the community early to maximize community engagement with the tools, and thus the data.

HLSPs & Catalogues

Many valuable community contributed high level science products (HLSPs) and catalogs are served to the community through MAST. Some of these result directly from data MAST holds, others are complementary datasets from other facilities. The presentations to the MUG highlighted recent developments in these areas and showcased some innovative new interfaces to access HLSPs. MAST plays a very important role in curating such datasets..

Finding: MAST has continued to improve the accessibility and organization of HLSPs, which is of great benefit to the community. This includes efforts to expand the availability and use of DOIs, providing guidelines for many different datasets to be ingested into the archive, and providing publicity and discoverability for the datasets.

HLSPs play an important role in accessibility of astronomical data. These science-ready products remove the barrier to entry in processing raw data, allowing immediate use for a

variety of applications. It is clear that these are of major benefit to the community, since the volume of data served by MAST in HLSPs rivals that of other MAST flagship missions. An example is the K2 HLSPs, which have been downloaded more often than the archived K2 data. MAST released 14 new HLSP datasets and updated 8 existing datasets this year.

Finding: Over the past year, MAST has made important progress in improving the automation of HLSP ingestion of observational data (time-series, images, spectra). For time-series data, in particular the process is now in an advanced state and has been applied to K2 and TESS observations. Efforts are in progress to improve ingestion of imaging and spectroscopic datasets as well, particularly in advance of the ULYSSES program. The best approach to dealing with catalog data, however, is not as clear yet.

Finding: Ingesting older HLSP data into the new system will require some effort, particularly in attaching appropriate metadata and updating documentation.

Recommendation: The MUG agrees that it would be beneficial to the community to devote the resources needed to improve HLSP automation, both for new coming datasets and for older datasets that need to be “re-ingested” with modern metadata and documentation.

One issue with HLSPs is discoverability. The datasets can also be heterogeneous, which makes standardized discovery methods challenging. MAST has been working on new interfaces targeted at specific communities for HLSP discoverability. One example is the prototype z.MAST which currently includes datasets derived from the CANDELS and GOODS samples. This is searchable through the web interface, an API, and astroquery. Another is exo.MAST, which centralizes HLSPs related to exoplanets.

Finding: The exo.MAST and z.MAST interfaces are a valuable means of enhancing the discoverability of various HLSPs, by targeting specific topical areas with a dedicated interface.

Recommendation: The MUG agrees that continued support and development of topical area specific HLSP discoverability portals would be valuable. MAST is encouraged to ensure these efforts usefully mesh with other similar archives like NED and NExSCi.

MAST currently serves a wide variety of catalogs, including Pan-STARRS 1, the Hubble Source Catalog, TESS and Kepler Input Catalogs, and a wide range of HLSP catalogs (like CLASH, PHAT, etc). These can be accessed in a variety of ways, including Casjobs, the MAST portal, APIs for scripted queries, and VO TAP queries. MAST has recently debuted the catalogs.mast interface, which links a UI + API and is easily tied to python notebooks.

Finding: The overall usage of PS1 and HSC catalogs are increasing, and users are accessing them through a catalogs.mast, VO TAP, and Casjobs interfaces, though the most frequent route for access depends on the catalog in question. Several other catalogs will soon be served through the catalogs.mast interface as well (TIC, KIC, GALEX, HLSP).

Recommendation: While the MUG recommends moving to TAP to remain aligned with trends in the field, MAST should monitor community response in the event a phasing out of casjobs has a negative impact.

Community interfaces

MAST Classic Redesign

The MUG was presented with updates to the MAST Classic interface. These tools are being modernised, yet still available for the current user base that relies on them. They are built for scalability, accessibility, and reliability, as well as reusability between missions, and are seen as a stepping stone towards a unified search system.

Finding: The MUG sees the MAST Classic redesign as an important step to maintain access to missions in a way that will satisfy the current user base. The ability to quickly transition from a set of text boxes to the equivalent API call was seen as an important strength of the new interface that will increase the accessibility and productivity of astronomical data.

The MUG was specifically asked to comment on the updated design, recommendations on rollout, and how to solicit user feedback. The MAST Classic tools are seen as an important service, and the analogies to the development of the new ADS are obvious. Preserving the ability to search with the classic 'feel' is useful, even as new backends or tools are developed around that interface.

Recommendation: Development of MAST Classic tools and their interface with the API should be continued, including on the future Unified Search system.

Recommendation: The MUG recommends soliciting user feedback from the community through demos made available at conferences such as the AAS meeting, in order to ensure a wide cross-section of the community can test the tools and discuss them with MAST members. The MUG also recommends rolling out the updated forms with a set of tutorials, possibly as YouTube videos or equivalent, discussing the motivation for the change and demonstrating how to use the new system, to help the community not only understand what has changed from the old form but why these changes are necessary and beneficial, rather than change for change's sake. When these are rolled out, a clear timeline should be publicized for the community to understand when the old pages will be retired.

MAST Website

The MUG was presented with an update on the MAST website. The team was motivated in their work by the feedback from the MUG's 2019 report, which the MUG commends. Specifically, the MAST team performed research towards the user experience, observing a sample of users

performing different tasks on the MAST website. This test taught the team some valuable lessons, and has led to some improvements and more in progress. The team has updated many pages on the website, with more in progress.

Finding: The efforts to improve the user experience on the website are appreciated, and the MUG is very enthusiastic about the work done by this team over the past year. This work is a highlight of MAST's efforts since the previous meeting.

Recommendation: The website team should develop a detailed plan and timeline to continue to incorporate the feedback from this test over the coming year, and continue the migration of all pages on the MAST website into the new format. The team should also discuss their usability research work with the rest of MAST (either through a seminar, an internal white paper, or similar) so the methods used to gain the insights they gained through these user tests can be applied to other facets of the MAST interface undergoing future (re)designs.

JDaVis

JDaVis, or Jupyter Data Analysis Tools, is a tool to be run in Jupyter notebooks, desktop applications and embedded in websites. Among other things, it is built off Glue, providing linking to astronomical data sets, and is agnostic to mission specific spectroscopic data. It has pre-built tool for 1/2/3-D spectra (Specviz, CubeViz, and MosViz), and has a customizable interface and configuration.

Finding: The development of JDaVis is motivated by the JWST spectroscopic initiative described in the MUG 2019 report. The MUG feels that by implementing this tool, it complements/augments the data visualization capabilities already possible. MAST's demonstration to the MUG on JDaVis's functionality and features were well received, noting that the presentation showed the value of video tutorials on the usage of tools developed by MAST.

Recommendation: The MUG encourages continued development of the spectroscopic tool. Keeping in mind the scalability of the project as well as the compatibility of it to other data analysis tools hosted by MAST (e.g., exo.MAST) and it's integration to the portal.

MAST Portal

The Portal was MAST's first integrated, multi-mission visual interface for data search and retrieval. Its rich feature set includes tools for searching, visualization, selecting, and retrieving science and ancillary data products. The Portal is a web application (written largely in javascript) that calls back-end MAST web services to provide most of the functionality. The python astroquery.mastlibrary uses many of these same web services. The Portal was intended as a replacement for the myriad of mission-specific custom interfaces, which use 2000's-era web technologies. The Portal supports most active and Legacy MAST missions, plus several selected catalogs. Data from future missions, such as JWST and Roman, and a handful of other projects, will only be accessible from the portal (i.e., they will not have the classic UIs).

Finding: The Portal is the main hub for data access to all MAST missions. The search function is very versatile, data may be visualized within the application, and easily downloaded (in several ways).

DOIs

MAST has undertaken an initiative to have authors publish papers with DOIs. The motivations include 1) papers with data links are more highly cited, 1a) data links in papers get broken over time, 2) enhances telescope bibliographies, 3) enables easy extraction of data from tables in papers 4) other places have done similar with success. The primary obstacle to date has been user adoption. One solution is the MARC project, which connects AAS Journals to Mast (and beyond, for example, IRSA, CDS, and other journals).

The Initial construction of DOI construction has been tested with *stsci.edu authors. This resulted in ~80% uptick among stsci authors, where DOI links were shown to have more hits than the ADS links. The plan moving forward is to roll out this functionality with 18 more institutions. This will include a question upon AAS Journal paper submission “do you want to contribute data created or re-processed for this paper as a HLSP?”. STScI will use the “ALMA approach” for JWST papers and have authors claim the DOIs upon ArXiv submissions (for ALMA, this has a 95% uptake on their identifiers). DOIs will be easily created through the MAST portal (and perhaps through the API and astroquery), and the generated DOI links will return the selected observations in the MAST portal.

Finding: The DOI system is now publicly accessible, but not very visible. Additionally, community uptake remains an issue. Expanding the test to the 18 institutions mentioned above will provide a better understanding on how to deal with user acceptance to creating DOIs for data in their papers. The MUG agrees that improved documentation and a tutorial will be well received by the community and help make this initiative successful

Recommendation. The MUG encourages MAST to explore DOIs to include enriched metadata, related datasets, and tools (notebooks, data documentation, etc).

Mission Closeouts

Missions end, but their potential for continuing science impact does not, depending on how well both the data and the knowledge of the mission behind it are preserved and curated. The MUG was presented with an example of MAST’s role in mission closeout with the closeout of the Kepler mission, which formally “ended” in fall of 2020. Of particular note in the presentation was timescale: *Kepler* closeout activities took many years, and lasted a significant time after the spacecraft mission was ended. A successful closeout, at least in terms of continued community use, requires significant overlap across the mission teams and NASA HQ. As presented by the

MAST team, the critical question becomes “how can people start from scratch 5/10/15 years from now, using only material and information in the archives?”. In the case of *Kepler*, mission closeout was significantly aided by access to the brain trust that created the mission, the data pipelines, and the light curve analysis software. After significant effort, the MAST and *Kepler* teams created a robust offering for the mission (https://archive.stsci.edu/kepler/) helping ensure continued community use of *Kepler* data for many years.
Finding: The MAST team has provided the community with a robust set of data products, documentation, analysis tools, and use tutorials.

Finding: The mission closeout efforts for *Kepler* are likely to ensure the continued scientific use of *Kepler* data for new and enhanced science for many years to come.

The MUG then heard a comparison between *Kepler* and *Galex* closeouts, and an important set of lessons learned from the *Kepler* closeout process. The lessons learned can be summarized as follows: missions must provide adequate resources for closeout, and do so before mission expertise is lost to inform it. The MUG strongly concurs.

Finding: A successful mission closeout requires dedicated resources, and early planning with archives.

Recommendation: MAST, as a condition for accepting the role of archive for a mission, should work with the mission to identify the resources needed for closeout, and set a work schedule to maximize archival utility of the data for futures where only the archive will be available for mission knowledge.

Recommendation: MAST, in conjunction with STScI leadership and NASA HQ, should develop guidelines for mission closeouts so that they may become part of the mission selection and development process.

The MUG was presented with a number of considerations relevant to the closeout of HST. By virtue of *Hubble’s* age, complexity, and vast user base, closeout for this mission will be significantly more resource intensive than that for *Kepler*. Although some closeout work has already begun, “Brain drain” is already occurring for *Hubble*, and the scale of the planning and implementation work exceeds the available resources currently at MAST.

Finding: Although work has begun on the closeout of *Hubble*, significant planning resources are needed quickly to avoid significant loss of knowledge needed to maximize the archiving and curation of *Hubble* data at MAST.

Recommendation: MAST, in conjunction with STScI leadership and NASA HQ, should convene a closeout committee to identify and prioritize the multiple aspects of *Hubble* closeout, and should begin allocation of additional resources to maximize archival use of *Hubble* data at MAST.

Looking forward, the suite of missions to be archived by MAST is vibrant and complex. Closeout activities for these missions are likely to be best accomplished if they begin now, even for missions at KDP-C or earlier, such as *Roman*. Likewise, missions with non-MAST archives should communicate, and ideally collaborate, with MAST to increase the utility of mission data going forward. In particular, the Great Observatories, given their wide community reach and impact, should be given significant consideration.

Finding: MAST, if properly resourced, is capable of executing closeout for current and future missions it will serve as the archive for.

Recommendation: MAST, in conjunction with STScI leadership and NASA HQ, should budget and plan for the closeout of complex missions such as *Webb* and *Roman* as early as possible.

Recommendation: MAST should consult with other archives on mission closeout activities, and consider collaboration to maximize science yield of the archives.

New Capabilities

MAST has developed prototypes of exploratory features that could further improve and enhance its search capabilities and presentation of its key datasets: including adding new metadata to provide to users and developing tools to serve previously underrepresented MAST user communities. This includes:

1) developing tools and web interfaces to sonify MAST light curves with the Astronify project

2) creating new utilities for moving object/Solar System body searches within MAST

3) the Hubble Image Similarity Project creating a large database of similarity information between *Hubble* images

Finding: The MUG commends the MAST team for its out of the box thinking; engaging with seed funding opportunities and community hack-a-thon events to develop potentially transformative small projects.

Recommendation: In this extremely challenging time due to the COVID-19 pandemic, we encourage the MAST team, when possible, to continue pursuing similar opportunities (both funding and community events/workshops) to kickstart experimental tool creation and produce small test-bed projects.

The MUG was provided with an update of MAST's exploration and development of possible moving object utilities. Solar System science was previously identified as an area for growth within MAST. MAST's primarily search tools and APIs are focused on cone searches and crossmatches with a particular right ascension/declination. Due to the nature of Solar System

moving objects, a single position search is not effective. Specific positions need to be searched for set time/date range. The MAST team presented two prototypes:

1) a query tool that searches selected archival datasets for potential serendipitous and targeted observations of Solar System objects based on JPL Horizons predicted ephemeris

2) pre-matched lookup for select Solar System objects tied to the MAST Portal search

Finding: The MUG finds that the prototype of moving objects search capabilities in MAST adds important value and utility.

Recommendation: MAST should continue to develop moving objects capabilities, and leverage the best elements of both prototypes going forward. The MUG recommends selecting the dynamic lookup for most Solar System queries and reserving the pre-match look up for a select set of Solar System bodies (the planets, space mission targets, giant planet satellites, dwarf planets, etc).

Recommendation: The MUG notes that the dynamic lookup has a potentially significant application in the Vera C. Rubin Observatory and the Legacy Survey of Space and Time (LSST) era, enabling precovery searches of Rubin Observatory small body discoveries in Pan-STARRS observations and other datasets hosted on MAST. Adding The capability to upload a predefined ephemeris or custom orbit would significantly enhance the dynamic lookup feature. The MUG recommends the MAST team explore adding this potential enhancement after the dynamic lookup search with JPL Horizons has been publicly released.

Recommendation: The MUG encourages the MAST team to liaise with the planetary astronomy community to promote these new features while still in development and when launched. Having a presence at the American Astronomical Society's Division for Planetary Sciences meetings and the Rubin Observatory LSST Science Collaboration sprints may be excellent avenues for scientific community outreach.

The MAST team gave an overview of their sonification project. A core mission of MAST is to make astronomy accessible to everyone, and this project was aimed at enhancing the experience of the Blind – Visually Impaired (BVI) community and identifying alternate ways of presenting light curves and time series MAST data. The COVID-19 pandemic has prevented in-class/public demonstrations, in-person testing, but despite these challenges the MAST team has created the Astronify[1] open source python package and designed 3 interactive exhibits that are currently under construction.

Finding: The MUG finds that the sonification of MAST light curves has significant benefit to blind and visually impaired researchers and high public outreach value.

[1]See https://astronify.readthedocs.io/en/latest/ and https://github.com/spacetelescope/astronify

Recommendation: The MUG recommends the MAST team work to incorporate Astronify sonifications for *TESS* and *Kepler/K2* light curves into the MAST portal and identify other current and future datasets where sonifications would be beneficial.

Recommendation: The MAST team should liaise with the STScI's Communications and Outreach Team and develop a joint plan for the Communications and Outreach Team to promote the MAST sonifications and create associated class-room materials.

The MAST team has been exploring avenues for complex searches, such as finding images in the archive that most resemble a template image. The MUG was presented with preliminary results from the Hubble Image Similarity Project, where volunteers reviewed a set of *Hubble* images and selected those that matched best to a template image.

Finding: The MUG commends the MAST team on their community outreach with the Hubble Image Similarity Project. The MUG finds that the Hubble Image Similarity Project adds new capabilities to MAST, providing a new avenue for MAST users to find similar *HST* images. The assessments are combined to compute a similarity distance matrix between all the pairs of MAST *HST* images. This technique could be deployed on other MAST datasets.

Recommendation: The MUG recommends the MAST Search By Image functionality including the output of the Hubble Image Similarity Project, be rolled out for Hubble imagery. After promotion of this new functionality, the scientific community's adoption and use of this search by image feature should be examined and assessed before the MAST team expands the Hubble Image Similarity Project to future datasets on MAST.