Report of the MAST Users Group (MUG)

October 20, 2021

Members of the MUG:

Britt Lundgren	UNC Asheville
Ben Montet	University of New South Wales
Brian Nord	Fermilab/Kavli Institute for Cosmological Physics
John O'Meara*	W. M. Keck Observatory
Karin Sandstrom	U.C. San Diego
Meg Schwamb	Queen's University Belfast

*MUG Chair

*NB: In this report, all recommendations/findings/key points from the MUG will be found as **bolded text** within the document.*

Introduction

It is the charge of the MAST Users Group (MUG) to *provide essential user perspectives on archive operations and development, including suggesting priorities for short and long term operational and scientific enhancements to the archive*. In service to the charge, the MUG meets annually and receives presentations from STScI staff, including Institute and MAST leadership. In 2021, the MUG met remotely due to continuing COVID spread mitigation restrictions at STScI. New to 2021 was a change in format to a shorter meeting, along with significant amounts of live demonstration of new/evolving tools at MAST. **The MUG strongly approves of the change in format, and recommends it for future meetings, but MAST should consider more frequent (biannual) meetings to allow for adequate time to be given to all major projects at MAST.**

The remainder of this document will follow the order of the topics as presented in the October 2021 meeting.

MAST in the cloud

The mission statement of MAST is to "maximize the scientific accessibility and productivity of astronomical data." To best achieve this goal over the next decade, MAST has identified the use of cloud technologies as a preferred path forward. The team has highlighted several key

benefits to this approach, including improved efficiency, opportunities to position MAST, and STScI more broadly in a leadership role in the astroinformatics space, and increased security. The team also highlighted how an increased presence in the cloud can be accomplished invisibly to the end-user in the access of MAST data, enabling user workflows to be unchanged by the change of backend, while also allowing for more visible cloud access for the users who would benefit from working in the cloud.

A significant benefit of the cloud approach is the ability to work closely with other archives (see *Fornax* below as well*)*, such as ESDIS, to leverage the lessons learned from those archives. This is likely to lead to additional exchange of ideas and methodology, and is seen as a strength of the MAST approach.

Finding: Through an analysis of user statistics, MAST has clearly demonstrated a community desire for cloud access to large datasets. For future missions, such as Roman, this need will grow significantly. The MUG finds that MAST's plan is carefully thought out, with many risks of reliance on the cloud addressed and considered. Continued care should be taken in future, measured steps.

Recommendation: The MUG recommends the continued development of tools and strategies to invisibly host data on the cloud, similar to the approaches currently undertaken with the TESScut software. The development of beginner and intermediate-level notebooks, providing examples of how to more effectively interact with these data than can be done locally and made visible and advertised broadly to the astrophysics community, would also significantly enhance these tools.

Recommendation: The MUG recommends that as more MAST initiatives include a cloud component, focus be given on community outreach education, as it represents a shift in how astronomers 'access' and 'use' data. The MUG recommends leveraging the AAS meeting opportunities (splinter sessions, hack-a-thons, etc) to assist in this focus on community education.

Documentation

MAST is in the midst of a 5-year long effort to migrate documentation accumulated over the last two decades to a modern publication platform. This includes more than 40,000 pages in a variety of early-2000s formats, some with stale or obsolete content. The MUG heard a presentation on the status of this migration and received a tour of the new documentation. The modern documentation will be published using two platforms: *jahia* which is what the main MAST homepage uses and *confluence* which powers MASTDOC and MASTDATA. The *jahia* site is useful for an overview of various topics, with a nice user interface that is uniform across the Institute's webpages. The *confluence* site is more powerful, with richer content creation tools and the ability for internal peer-review of material. The documentation on *confluence* is cross-linked extensively between pages to keep material from being repeated multiple times in different places. The MUG were given a tour of the new system, demonstrating the backend for

the drafting of pages, the process for internal peer review, usage statistics, organization of menus, and more.

Finding: The migration and updating of documentation is of great value to the community. MAST is to be commended in the care that is being put into making this material easily accessible through these new publication platforms.

Recommendation: The MUG recommends involving end users in testing the interface early in the process to incorporate feedback from the community. Testing the experience of a wide range of users may reveal issues that otherwise may not be noticed till the pages are fully migrated. The MUG recommends building on the methodology developed and lessons learned from the MAST website redesign from last year, as detailed in their 2020 report.

Recommendation: MAST should do community outreach to demonstrate the new documentation repository at the AAS meeting and record outreach sessions for later access by the wider community.

Recommendation: The MUG suggests some further exploration of how the new documentation content and organization can connect with Jupyter notebooks, which are becoming increasingly important in many aspects of MAST-related data analysis.

Moving Targets in MAST

The MAST team has made significant progress on developing Solar System moving object tools. The MUG was given a demonstration of how these tools worked with the Hubble and IUE data and examples for what the web interfaces would look like. Incorporating the last year's MUG feedback, the moving targets search design and development incorporates a hybrid approach. The search tool utilizes both a dynamic lookup based on the JPL (Jet Propulsion Laboratory) Horizons service and a pre-matched lookup for select Solar System objects. The combination of the ESAC (European Space Astronomy Centre)-style pre-matched search results with a CADC(Canadian Astronomy Data Centre)-style dynamic search.

Finding: The MUG commends the MAST team on the development and progress of the Solar System moving object tools. The MUG approves of the plan to utilize both a dynamic lookup based on the JPL Horizons service and a pre-matched lookup for select Solar System objects.

Recommendation: The MUG recommends continued development and roll out of the moving object search tools to serve the broader planetary community.

Recommendation: The MUG suggests exploring a time cutoff or positional error uncertainty ellipse cutoff for the results returned by default. If in the dynamic search, the Solar System small body has 1 month of observational arc then searching 20 years of Hubble observations may not

be productive as the error ellipse arc will be significantly bigger than the sky covered in individual images/instrument fields-of-view.

Recommendation: The MUG recommends incorporating the ability to search on orbital parameters for new discoveries not yet sent to or processed by the Minor Planet Center. We believe this will be an important feature for the UI, as the MAST moving objects tools could be used to search for pre-recovery observations to better refine the orbit of a new interesting Solar System discovery.

Recommendation: The MUG encourages further promotion of these (prototype) tools. In particular, the MUG encourages further engagement with the planetary astronomy community. One such avenue could be a booth/information session at the American Astronomical Society's Division for Planetary Sciences (DPS) meeting.

Fornax

The *Fornax* project aims to facilitate work with archival data across archives, specifically IRSA and HEASARC along with MAST. The MAST team meets regularly with representatives from the archives, NASA APD at HQ and Booz Allen. The goal of the project is cloud focused, with a suse case exemplar of forced photometry across data from multiple NASA missions: *GALEX, Spitzer, XMM-Newton,* and *Chandra*. Resources have been allocated and significant planning is in work for FY22 and 23. The MUG was excited by the opportunity that the *Fornax* project represents for realizing the larger goal of direct-to-science activities across multiple archives for the end user.

Finding: The MUG is excited by the opportunities *Fornax* presents for significant cross-archival science in the cloud.

Recommendation: The MUG encourages further development in *Fornax* and requests an update at the next meeting to assess progress and lessons learned from the cross-institutional collaboration.

Modernized Search Interface

The MUG received a presentation on a new search interface, with HST search as an example. The new interface (deployed at https://mast.stsci.edu/search/hst) was designed by a professional interface designer with a strong focus on accessibility. The MUG was extremely impressed by the new interface, including the use of descriptive text, video with text, real-time object name resolving and more. The API query generation was also very impressive.

Finding: The MUG was very impressed by and strongly supports the development efforts around the new search interface, particularly with the focus on accessibility.

The only concern expressed by the MUG on the interface is whether users will find it by default. A web search for ‘MAST HST Search’, for example, does not yield this new interface as the first link, and there is a concern that the community may not always find this otherwise excellently designed, accessible, usable tool.

Recommendation: The MUG recommends that MAST make efforts to make the modern search interface the default and most easily found option for users.

The MUG is excited to see the future evolution of the interface, including the cross-collection search via unified search.

TESS

The ongoing TESS mission provides one of the most popular datasets in the MAST holdings. As the TESS mission continues, additional data is delivered to MAST approximately fortnightly. The mission may produce more than 1 petabyte of data by its end, making it the first Pb-class NASA mission and producing a reasonable test for MAST as they prepare for future Roman data.

Over the past year, the MAST team has continued to develop tools to increase the accessibility of TESS data hosted on MAST. One of these is the “moving target” search on TESSCut that enables small cutouts of TESS data to be downloaded that track solar system objects. The team has also hosted new HLSPs of broad use to the TESS community, including the TESS Image CAlibrator (TICA) full-frame images which are delivered within one week of the end of sector observations. Over the next year, the team plans to move more of the TESS services to the cloud, with the intent that the user experience is identical to the current workflow.

The notebooks designed to provide an introduction to TESS data are useful and seen as a significant net benefit to the community; the outerspace site “Ways to search and interact with TESS data at MAST” and the GitHub TESS notebooks both provide a suitable introduction to the data; however, neither has been updated in more than one year and so the team should ensure these are regularly maintained to ensure public awareness and accessibility of the new tools.

Finding: The MAST team is doing a commendable job keeping up with the rush of TESS data and ensuring the data is maximally useful to the community on an efficient timescale. This work builds on their previous efforts with Kepler/K2 and serves as a stepping-stone to dealing with future Roman data. MAST is clearly providing a net benefit to the community through their TESS efforts, which are clearly in line with the MAST mission.

Recommendation: The MUG recommends continued development of tools to serve the broader TESS community, and the continued development of documentation to make these tools accessible. The MUG suggests that MAST may find benefits in reaching out to researchers using TESS data beyond the stellar and planetary community to understand their needs and

devote efforts towards ensuring those needs are being met, building upon their work on the moving target TESSCut tool.

Hubble Advanced Products

The Hubble Legacy Transition Project is a new effort to seamlessly integrate the high level science data found in the Hubble Legacy Archive (HLA) into MAST. This will involve the creation of Hubble Advanced Products, which are HLA-like data products generated by an automated pipeline that can be run on updated data after any reprocessing campaign. This will allow up-to-date advanced science-ready products including mosaics and source catalogs to be accessed through the standard MAST portal. An important element of this process is setting quality requirements on the products and communicating the data quality to the users.

The HLA components that will be integrated into the Hubble Advanced Products include improved astrometry calibration built on Gaia, single visit mosaics and catalogs, and multi-visit mosaics for all public ACS/WFC and WFC3 observations drizzled on predefined pixel grid. In addition this effort will provide user tools, particularly the functionality in *drizzlepac* which allows users to create their own mosaics for proprietary data or use custom pixel grids or create mosaics for other instruments (e.g. WFC3/UVIS). The new astrometric solutions are impressive, with 10 mas precision relative to Gaia, representing a major improvement over standard pipeline products.

Finding: The MUG was impressed with the quality of the HAP mosaics and astrometry solutions. It is clear that these advanced products will be highly valuable to the community.

Recommendation: The MUG recommends continued investment in the HAP project. Members of the MUG were interested in the interaction between Hubble imaging HLSPs delivered to MAST and the up-to-date HAPs. It would be worthwhile to consider the use cases for both of these datasets and whether some clarification for users would be worthwhile.

TIKE

Over the past five years Jupyter notebooks have proliferated throughout the astronomy community as a popular tool for writing and sharing code to analyze and visualize data. Cloud-based platforms utilizing Jupyter notebooks have already been developed and adopted by multiple communities of ground-based astronomers (e.g., NSF's NOIR Astro Data Lab and the Sloan Digital Sky Survey's partnership with SciServer.org). The primary advantages of these platforms are that they simplify access to and analysis of large datasets by bringing the user to the data (and not vice versa), and sidestep the need for any software or data downloads. Jupyter notebooks have also become a successful tool for launching students into research.

The Timeseries Integrated Knowledge Engine (TIKE) serves to bring this popular functionality to MAST. This newly developed cloud-based computing platform facilitates the exploration of

MAST data with Jupyter notebooks. The framework is freely accessible to anyone with a registered MAST account. While still in the early phases of production, the platform is fully functional, and a few example Jupyter notebooks have been developed to demonstrate its usage. In addition to providing a sandbox from which to easily query, analyze, and visualize MAST data in the cloud, TIKE offers limited storage space for users to upload additional data -- an important feature that enables cross-matching with non-MAST data. The Jupyter notebooks created in TIKE can also be readily shared and edited collaboratively by teams using the nbgitpuller routine.

Finding: The MUG was very excited about this new tool. The TIKE has the potential to meaningfully expand the accessibility - and therefore also the science output - of MAST data. This platform could also be used to develop high-impact educational activities for undergraduate and graduate students, leading to the TIKE having a very broad impact.

Recommendation: It will be helpful for users to have access to a larger and more diverse suite of example notebooks with simple, straightforward recipes for common data querying and analysis needs. Members of the MUG suggested MAST host a library of vetted notebooks that are created by MAST developers and/or contributed by other TIKE users. A workshop demonstrating this platform at a future AAS meeting would raise its visibility and adoption by the community. If possible, it would be helpful to also host a discussion forum for MAST users to share questions and help.

User Support

MAST holds a diverse set of data and resources within the archive. With the onset of large wide-field surveys, the need for archival data will continue to grow. Currently user support is focused around each tool with the developer or development team focusing on community engagement. As MAST continues to develop its portfolio of tools and services, the MUG sees a need for a unified user support model to be explored. This may potentially include developing a dedicated position that focuses on wider user support issues/planning.

Recommendation: The MUG encourages the MAST team to develop a long term user support strategy and examine how a unified user support and engagement strategy may be implemented as MAST continues to develop into the Fornax era.

Recommendation: As we enter the era of operations for the Vera C. Rubin Observatory, JWST, the Roman Telescope, and Euclid, the need for archival data will only increase. The MUG notes that a significant fraction of the future MAST user community will be looking for MAST datasets to supplement/combine with the data from these upcoming missions and surveys. MAST providing the raw data may not be enough to facilitate science for those users previously unfamiliar with MAST's datasets/resources. It would be helpful to explore tools/utilities for providing reduced observations as output from MAST's query tools such as the moving object search tools.