

Response to the 2013 MUG Report

We thank the members of the MAST Users Group (MUG) for their very helpful recommendations in the report that was submitted to MAST after the 2013 MUG meeting. We have considered the various recommendations in their report, and we are submitting this response to provide an update to the MUG on how we are responding to the recommendations in the report. Blue italic font is used to indicate items from the MUG report, followed by black font to indicate the MAST response.

Overview Comments

The MUG would like to see a slide presented at the onset of the MUG meetings outlining the structure of the Archival Branch, highlighting those individuals in charge of the interactions between the MAST and the community, and how the MAST personnel formally interact with other STScI branches.

We will include this information, either at the outset, and/or as part of the information that is provided prior to the meeting.

Carryover – WFC3 Persistence

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

MAST supports this recommendation, and our representatives have continued discussions with the WFC3 instrument team. Currently the WFC3 instrument team believes that the persistence correction is not sufficiently robust to include in a pipeline (they think it is not guaranteed to improve the data in all cases). MAST will

continue to advocate for the users of the archive on this issue, and we will work with the WFC3 instrument team to encourage either improvements to the algorithms or to identify a flexible approach to the pipeline processing (e.g., to provide products with and without persistence correction). In the meantime, MAST continues to provide access to the persistence corrected data products created by the WFC3 team including any updates and changed files. These data are online in an area where we can control access to the proprietary data.

Survey and Community Feedback

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

Feedback tools can be implemented through popup windows at critical times during data retrieval/access that query the user to report problems and feedback, as well as at the end of the entire portal session. One or more “Help”-type buttons could also be located in and remain on each page of the portal session, so if an issue materializes along the way, the user can click for instructions on how to proceed or report. Examples could be “Problem” and “Feedback” labels, which should be clear to most users. Depending on the nature of the problem, one or more of these features could redirect the user’s response to the normal helpdesk, and provide a priority flag. Such details in the implementation of the scheme can be worked out internally within MAST.

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

A portal ticket has been filed to add labels such as “problem” or “feedback”. STScI is looking at the possibility of “forum” software and the archive has expressed our interest in this topic. We anticipate that once the single sign-on is implemented we can begin to use this forum software. We are not convinced that pop-up exit surveys will provide sufficiently valuable information to justify the accompanying user

annoyance (how many of those surveys do you fill out yourself?), but we'll give that approach further consideration as well.

MAST Portal

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

MAST is beginning to work with a graphic designer from OPO and will bring these suggestions from the MUG into the discussion. Adding the news feed to the front page of the Portal has been added as a ticket, and how it would be most effectively integrated into the Portal will be added to the discussion with the designer.

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

The portion of the MUG meeting concerning the portal was a highly interactive phase of the meeting, and some points about our release plans were obviously not communicated well. MAST has always planned to release the portal in phases and


not as a one step replacement for existing features. MAST felt that the new features available in the portal were of significant enough value to release, with features from existing interfaces and planned enhancements phased in over time. We always intended to do the first portal release in November 2013, and never stated otherwise. This includes new capabilities that will be possible after the HST upgrade project and creation of the HST online cache are released. We apologize for leaving the MUG confused about the schedule and plans. We will ensure that future presentations highlight that sort of information earlier in the talk before embarking on interactive demos.

Additional Comments from the MUG on the MAST Portal

Tutorials are very helpful. They should be presented as screen shots on a help page for the new portal and/or be video (webcast) presentations highlighting use of the Portal, its sub-domains and tools accessed by a link or button.

MAST is currently starting to look at all documentation and this recommendation will be considered in this project.

All classic features of the current MAST data access pages need to be carried over and immediately implemented into the new MAST Portal. This includes features like viewing spectra available for download. Thereafter these features can be extended for the newer MAST holdings.

All spectral preview and viewing options available in MAST classic are also available after clicking on the Information button. 

All of the table columns and spectral features for all holdings need to be completely defined for the user via pull down information or linked pages describing these. Units need to be clearly listed. Computing of magnitudes and the magnitude systems used should be clarified.

Pan-STARRS is a great addition to the MAST holdings. While it is understood that separate hardware exists for processing, access (downloading) of the data will occur through the present bandwidth available through MAST. If this includes images, it may over burden the MAST network and slow down data access for other MAST users. For such large time- domain datasets like Pan-STARRS, is it possible to secure separate fiber-cable links to access these data?

STScI does not have the resources to set up separate internet connections for different projects. If the bandwidth usage by Pan-STARRS (or any other individual project) becomes excessive, we will need to deal with that through a combination of upgrading our network connectivity for all projects and by imposing restrictions for data retrievals. We have faced this sort of issue in the past during major public data releases, massive retrievals of DSS images, and (most recently) the Comet ISON

episode, so we do have relevant experience. However, we anticipate that most of the challenges for Pan-STARRS will be associated with querying the very large, complex database rather than downloading large data volumes; that database load will be well managed and segregated from other missions by placing it on separate hardware.

All spectra should be viewable prior to download. GALEX and slit-less spectra need to be included in this fashion.

Extracted GALEX spectra do have previews available via the Portal. Extracted slit-less HST spectra do have previews available, but most of this data remains unextracted.

All proposal target name formats should be searchable within the Portal. Further, discussion should be held between MAST and the Proposal development team to allow standardized (i.e., SIMBAD, NED) target name formats to be accommodated. MAST targets should be transparent and searchable under these names. As an example of the type of problems alluded to here are some of the planetary nebulae targets. HST proposal formats have not allowed in the past formal names such as PN Gxxx.x-xx.x. Proposers are forced to be creative, renaming their targets, such as PN-Gxxx.x-xx.x, which are not searchable under the beta-portal. All MAST targets should be cross-correlated with their scientific names, and perhaps having these listed in a pull-down or pop-up window is possible.

Users may search the Portal using any name that is resolvable by the SIMBAD or NED services. The actual searches are done using coordinates, so the name the original GO gave it does not have to match the resolved name exactly. However, we do understand the frustration that users feel about this matter, and future implementation of string searches or filters would be facilitated by more uniform names. The archive has been in communication with the APT team asking for this and a little progress has been made. For JWST the users will have the opportunity to include a resolvable name in addition to the name given by the team.

MAST can consider the cross-correlation suggested above, but may not have the resources to do this and keep the selection of names up-to-date, as many targets have several names, and the names and the formats are in a constant state of flux. MAST would prefer to put resources into better tagging the targets with categories (e.g. quasar, interacting galaxy, etc).

Standardizing some features of the portal with those used daily on the internet is helpful (i.e., common terminology), perhaps changing the shopping basket to a shopping cart. Old items retained in the shopping cart should be differentiated from new items being added in the current session, using colors as an example.

The HLA uses a shopping cart icon / terminology. During HLA feedback activities at the AAS and other meetings, many users assumed there was a cost to retrieve the data. A basket was then chosen, hoping that the assumption of cost might be avoided.

Suggestions about selection content and retaining and identifying contents in the basket will be considered during rework of this function, which already has a ticket.

The archive should give descriptions of the formats of the different datasets. This is especially helpful with spectra, whose formats vary a lot from one instrument to another.

Agreed. We will add this as a ticket.

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

This suggestion will be added to the list of items to be considered as part of a documentation upgrade.

High Level Science Products (HLSP)

Types of high-level products that could be of wide interest to the community are often produced through NASA and NSF funded programs. With NASA these are primarily associated with large observing programs obtained through the individual mission calls. Closer involvement with MAST during the selection process and performance periods to ensure these products reach the widest possible audience would benefit the community at large, and undoubtedly the teams preparing these datasets. Early decisions to include (or not) relevant data products and tools within MAST will help in disseminating the products much earlier and in ensuring that they are useable within the construct of MAST. As regards HST and JWST, MAST should be involved in the process at all stages and be in contact with the PI after selection.

For several years, once the announcement of accepted proposals has been made, MAST archive scientists have written to the PIs of Treasury, Archival Legacy and Large programs, inviting early interaction for preparation and delivery of HLSP. For the Treasury and Archival Legacy programs, periodic (~ six months) mail is sent asking if we can be of assistance. All mail refers the PIs to the guidelines which are posted online. Once the program PIs are ready to begin delivering data, MAST staff then coordinate discussions with them on the details of what types of products they are delivering, as well as the timescales (especially helpful for extended programs or cases with multiple releases). MAST staff also then work closely with the teams once they begin to deliver data, to ensure that the products satisfy the MAST delivery guidelines, as well as setting up appropriate web pages for public distribution.

Liaise with ADS (and with AJ, ApJ, PASP, MNRAS, and A&A) for tables, spectra, and images of high-level science products related to MAST holdings. Standards need to be defined and assured in the data products.

MAST will consider the feasibility of this although execution will be dependent on available resources. We are having on-going discussions with ADS and the other NASA archive centers (HEASARC, IPAC, NED) about how to better integrate MAST data with publications. Josh Peek is leading a joint MAST-AAS project to explore new ways to make this more effective. He will report on this at the MUG meeting.

*Liaise with ADS about putting direct links to HLSPs in their "On-Line Data" links. This is done with Vizier tables. As it is now, "On-Line Data" has a *general* link to MAST, followed by a link to the low-level data products in MAST. And when HLSPs exist, there should be a direct link to them.*

MAST has done this. Initially, there is a link to the HLSP page for those papers identified as the seminal paper preparing the HLSP. Future enhancement, pending availability of resources, will link to the data for those papers that have utilized the HLSP.

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

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

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

We have sent announcements to mailing distribution, announced new HLSP on our newsfeed, twitter and facebook accounts. An article about MAST HLSP is part of the AstroBetter blog. In addition, MAST contributes an article to each STScI Newsletter containing announcements of new HLSP products, including a description of the various products that have been recently delivered by each team, as well as a discussion of how they might be useful for science done by others in the community.

Kepler

We agree with the MAST group's view that stitching together different quarters of data is still a research problem and there is no accepted way of doing so. However, python scripts already exist for stitching together the different quarters and we urge

MAST to make these available with documentation. Nevertheless, the added science value of having a reliable means of stitching together different quarters of Kepler data is huge. While we understand that the MAST team should not be responsible for tackling this, and that the Kepler planet-finding team has no intention of pursuing this important endeavor, we encourage MAST to liaison with the Kepler Asteroseismic Science Consortium in providing these resources when available. In fact, high-level data products derived from this consortium effort should be ingested into MAST in a coherent and timely basis.

We have provided links to science analysis software through the left-hand “gutter” on any of MAST’s Kepler pages. We have recently re-named and re-organized these. You can now visit the Kepler Science Office’s own page containing a nice collection of software links for analyzing and viewing Kepler data, including PyKE. Simply click on “Data Reduction & Analysis” on the left-hand gutter of any MAST Kepler page, where users will now find plenty of links to third-party software.

We have contacted the Kepler Asteroseismic Consortium and asked if they would be interested in archiving any of their data products at MAST, including the working-group-led de-trended and stitched light curves, but the KASOC said they were not interested in doing so.

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

The name-resolver has been updated to understand Kepler IDs, and can now be used in the Portal (and in fact any other service at MAST that uses the name resolver).

MAST should liaise with the Kepler Asteroseismic Science Consortium to ingest higher-level data such as “corrected” time series suitable for asteroseismic analyses and power spectra.

As previously noted, we have contacted the KASOC, and they are not interested in providing their data for public access at MAST.

MAST should contact authors of Kepler-data related papers for higher-level products such as masses, radii, stellar ages, and such. These higher-level products are currently digested automatically by CDS and appear as a link in ADS. It would be useful if these data products were also available on MAST, though this should not place a burden on MAST in doing so. Perhaps providing links within the data portal is the best approach here.

We will continue efforts to reach out to authors of Kepler-related publications with substantial data that could be used as HLSP, however, we note that some of these

are already archived in one form or another through other sites (KASOC, CFOP), and in those cases it is expected authors will be even less willing to spend the effort to prep their data for ingestion into MAST. Increasing the linkage between known Kepler papers and the Portal is on our to-do list for the future. In point of fact, this issue is applicable to all MAST missions in the Portal, and is not limited to Kepler.

A somewhat improved version of the Kepler Input Catalog will be published soon. Mast should contact Daniel Huber, the lead author of the paper, for the new catalogue.

This catalog has been released, and MAST gets a copy of the data via NExSci. There are database tables with MAST's Kepler CasJobs pages for users to query already. We have developed "MAST-classic" search interface for these tables, and are in the process of integrating the Q0-Q16 stellar table from the Stellar Parameter Working Group with our two main database tables used in Target Search and Data Search, respectively.

GALEX

Since the original GALEX teams observing in the scan mode have not been successful in scientifically reducing and analyzing these data, it is the MUGs view that besides storing these data for public download, nothing else should be done with them inside MAST at this time. MAST has other more important immediate tasks to deal with.

We are beginning to post scan-mode data, as delivered, in a publically accessible html/ftp location, with basic documentation. Aside from the natural desire to use MAST's gPhoton tool with these data, MAST does not envision doing any more activities with the scan mode data once all of it is available to the public for download.

Provide a link to or insert within MAST a detailed description of the data reduction methods employed for the GALEX archived data.

Transferring all externally hosted documentation that MAST references to internal pages is an action item for MAST in the coming year.