

NASA Data Center Annual Program Plan
NASA Grant Number [NNX13AC07G](#) / STScI Grant J1561
January 2013 through July 2013
Barbara A. Mikulski Archive for Space Telescopes (MAST)
(Optical/UV Science Archive Research Center)
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218

Overall Mission: MAST supports active and legacy mission datasets and related catalogs and surveys, focusing primarily on data in the ultraviolet, optical, and near-IR spectral regions. Support includes providing data curation, providing expert support to users of the data, providing access to data-specific calibration and analysis software, providing user support for this software, and maintaining public access interfaces to the data. This report covers data financially supported under the MAST contract. Archive and distribution activities for HST data are supported under the HST contract; the Kepler contract supports some of the archive activities for Kepler data. Any JWST archive activity is financially supported under the JWST contract. Some of the statistics include HST, Kepler, and JWST data volume and usage statistics. Some projects are funded from all archive funding streams, including the MAST grant.

Holdings and distribution

As of August 1, 2013 MAST holdings are over 251 TB, including 70 TB of Hubble Legacy Archive data, nearly 77 TB of HST standard pipeline products and nearly 4 TB of Community Contributed High Level Science Products. Figure 1 shows MAST holdings as of July 1, 2013.

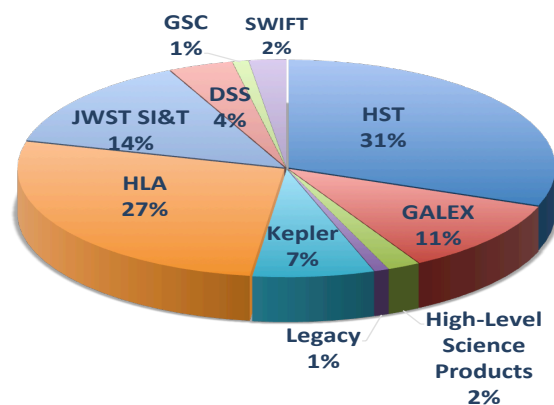


Figure 1: MAST Holdings

The archives at STScI have distributed over 130 TB of data between Jan 2013 and the end of July 2013, and have ingested nearly 17 TB of data. The figure below shows the statistics on data ingest and distribution to the public from Jan 1995 through July 2013.

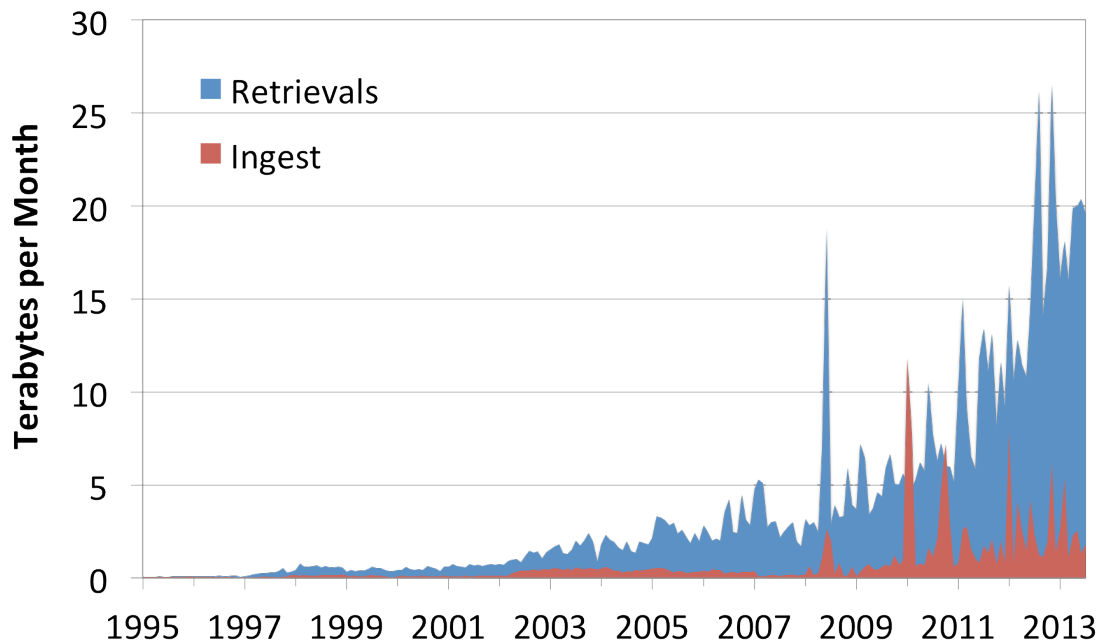


Figure 2: MAST Ingest and Distribution

Mission/Project reports

Hubble Legacy Archive (HLA) and Hubble Source Catalog (HSC)

On May 8, 2013 the HLA Data Release 7.1 (DR7.1) was delivered to the public. Much of this work was to support the Hubble Source Catalog Project (HSC). The HST Mission and MAST support the catalog development with interfaces integrated into the HLA and standalone interfaces also developed by MAST. The HSC contains information about sources detected in many of the ACS and WFPC2 images. It is based on the positional crossmatching of source lists in the HLA across HST visits (pointings) and/or detectors. The HLA interactive display now includes HSC catalog overlay and other advanced control options. Work is continuing on improvements to the catalog, with version 1 planned for release later in 2013.

Galaxy Explorer (GALEX)

Mission Conclusion:

After 10 years, the GALEX spacecraft mission successfully ended when the spacecraft was turned off on June 28, 2013. Although the spacecraft itself is will no longer be collecting data, a variety of additional data products and services are planned at MAST as we transition to the GALEX legacy period.

Data Ingests:

The final GALEX data release was the so-called GR7, which was made publically available at MAST in Feb. 2013. The GR7 largely consisted of observations from Feb. 2010 to Feb. 2012, after which the spacecraft was lent to CalTech to operate on private funding. A total of 9,906 new tiles were released as co-adds, along with 1,105 tiles that were re-released due to additional observations taken over the GR7 timeframe. All of these data have been integrated with our services, including CasJobs, GalexView, and our Search Forms.

gPhoton:

MAST has been working with a former member of the GALEX pipeline team to create a database of every photon observed by the spacecraft. This large database would allow

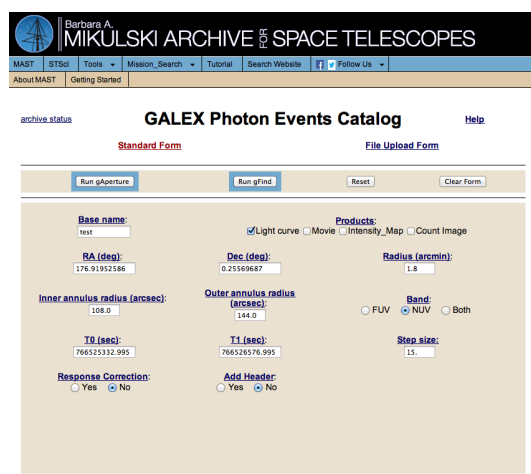


Figure 3: gPhoton web form (early alpha version).

for the creation of time-series lightcurves of arbitrary time sampling, improved background correction, animated movies to identify variables, and intensity maps. The associated analysis tool, called gPhoton, is a Python-based command-line software package that allows users to find GALEX data given a coordinate on the sky and generate lightcurves, intensity maps, and movies. An “open beta” was begun in early summer with a handful of interested external members, and a stable version of the command-line gPhoton software package was released to these early adopters. We continue to make updates and improvements to the software package. In addition, an “alpha” version of a web interface was developed and tested that will eventually let MAST users generate these data products through a familiar search form interface (Fig. 3). One notable error in the GALEX database that was fixed by debugging of gPhoton was a small wobble in source coordinates. We were able to identify the source of this wobble and correct it, which will result in (relatively small) improved astrometric solutions once applied to the GALEX data in its entirety.

for the creation of time-series lightcurves of arbitrary time sampling, improved background correction, animated movies to identify variables, and intensity maps. The associated analysis tool, called gPhoton, is a Python-based command-line software package that allows users to find GALEX data given a coordinate on the sky and generate lightcurves, intensity maps, and movies. An “open beta” was begun in early summer with a handful of interested external members, and a stable version of the command-line gPhoton software package was released to these early adopters. We continue to make updates and improvements to the software package. In addition, an “alpha”

GR6 and Kepler GCAT's:

Seibert et al. have created unique catalogs of GALEX sources (GCAT's), which simplify a lot of analysis by removing the multiplicity of sources in the (GALEX pipeline's) MCAT catalogs. Two versions of a GCAT are created: one using the broad (but shallow)

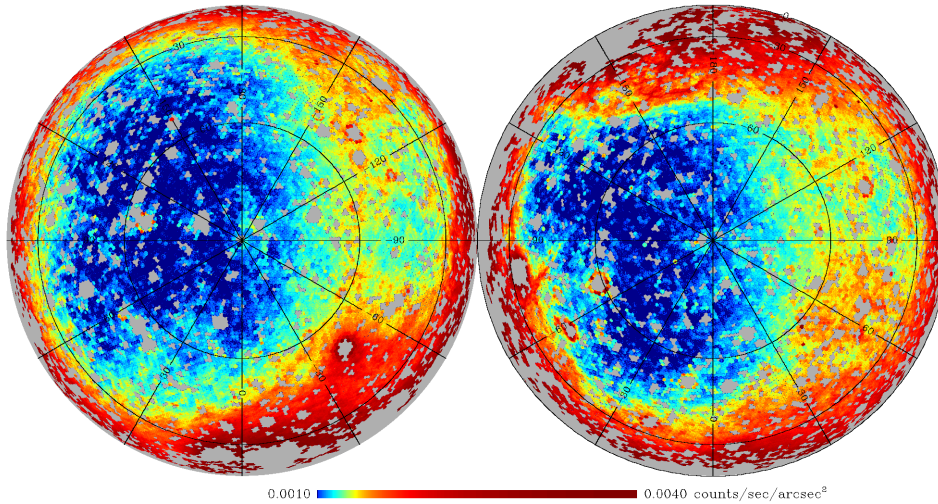


Figure 4: Sky coverage of GR6 GCAT (AIS)

AIS data (Fig. 4), and one using the deeper (but with less sky coverage) MIS data. Seibert et al. created a GCAT using the data up through GR6 (but not including GR7). There are no plans to create a GCAT using the entire GR7 database, but the GR6 GCAT contains nearly 40 million unique sources from the AIS and over 22.5 million unique sources from the MIS. Each source has multiple flux measurements using different aperture sizes, and the GCAT team has improved pixel artifact removal by visually inspecting every GALEX tile.

Seibert et al. have also created a special Kepler GCAT using all data through the GR7 data release that overlaps with the Kepler field of view. Although the same software was used to create the Kepler GCAT as was used to create the broader GR6 GCAT, there are some differences in the implementation that will be documented at MAST upon release. The Kepler GCAT consists of over 550,000 sources that will be a great asset to MAST users querying the Kepler field for data, making it easy to obtain UV fluxes for Kepler targets of interest. MAST plans to integrate and release both the GR6 GCAT and the Kepler GCAT over the next couple months.

Post-NASA AIS Data:

We now have the post-NASA (CalTech-led) AIS data at MAST, which consists of galactic plane scan-mode data. We intend to release this data in the future, but are working on determining the best way to integrate these data with our current holdings.

GCAT was created are a bit different from the GR6 GCAT, but they are largely similar to the broader GR6 GCAT. The Kepler GCAT consists of 551,479 sources that will be a great asset to MAST users querying the Kepler field for data, making it easy to obtain UV fluxes for Kepler targets of interest. MAST plans to integrate and release both the GR6 GCAT and the Kepler GCAT by the end of 2013.

Kepler

The Kepler Data Management Center (DMC), funded directly from the Kepler Project, was established at STScI to archive the Kepler Data products. Kepler project team members, Kepler DMC staff, and MAST staff members continued to collaborate to design and implement several search interfaces and website documentation and content.

Data Ingests:

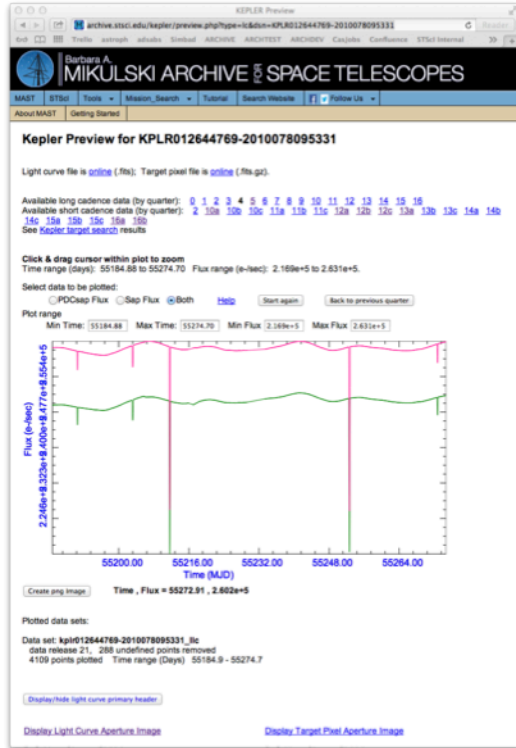
The Kepler team released several new quarters of data to MAST. Quarters 14, 15, and 16 were released in Feb., May, and July, respectively. In addition, the Kepler team re-processed all previous Quarters (Q0 – Q14) with an updated pipeline, which featured improved lightcurve detrending and corrected timestamps in all the lightcurve files.

From external contributors, we ingested the second half of the Kepler INT Survey (KIS). The KIS Part 2 more than doubled the KIS coverage to ~97% of the Kepler field, containing a total of over 14 million rows. In addition to the previous broadband filters, MAST also made the narrow-band H-alpha fluxes from the KIS survey available. All of these fluxes were cross-matched to Kepler targets, and these fluxes are now available through the Kepler Target Search Form and through CasJobs.

Several database tables are now retrieved from NExScI and made available at MAST. These include the KOI table, Kepler Planets tables, and NExScI Target Name Relations table. We are working on how best to include the Stellar table in our holdings in the near future. In addition, we continue to update the eclipsing binary and known red giant tables on a regular basis.

Lightcurve Preview Pages:

After performing a Query on our Kepler Data Search Form, users may examine the lightcurve files via an interactive plot tool (Fig. 5). These “Preview Pages” contain a wealth of additional information, and we have expanded their content and functionality over the past several months. Users can now jump between different data sets for the same target using the links at the top of a Preview page. They can quickly download a lightcurve or Target Pixel File for the plotted data set on these pages, and see the lightcurve and target pixel aperture images. In the near future, we plan on adding additional data and features to these Preview pages, including automatic loading of the aperture images on the page and a wealth of variability statistics.



During the next year, we plan to integrate existing functionality for exploring the Kepler Full Frame Images (FFI) with the interactive display tool, developed for the HLA, into the MAST portal interface.

Figure 5: Example Kepler Preview Page

Hopkins Ultraviolet Telescope (HUT)

New Data Products:

The HUT team re-delivered the Hopkins Ultraviolet Telescope (HUT) data to MAST using a new pipeline. These data were made available to the public in March. The data format was changed to make it much more user-friendly, including FUSE-style time-tagged photon event lists, and extracted and calibrated spectra, both of which are now FITS-standard compliant. Quick-look plots that were much easier to read compared to the previous data release were added for every spectrum (see Fig. 6), and include the extracted spectrum, photon count rate as a function of time, and pointing error diagnostics during the integration. For the first time, the TV guider images were included as downloadable files so that users can visually confirm the orientation of the spectrograph slits relative to the source. A complete re-write of the HUT project pages at MAST to reflect these changes was done, and all

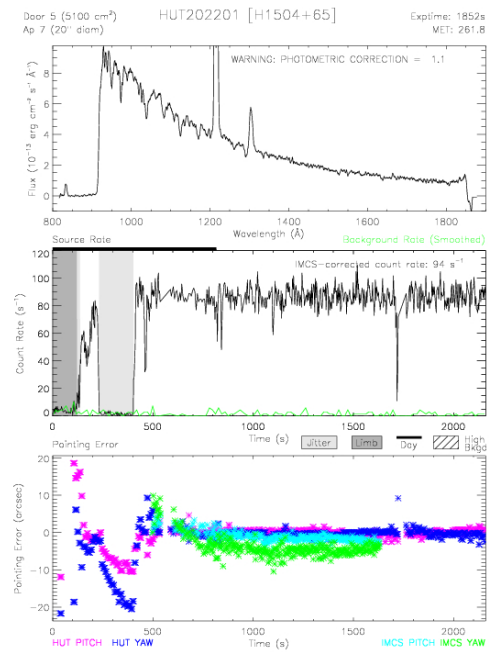


Figure 6: Example of the new quick-look HUT plots.

the new data were made available through MAST's search interfaces in time for the publication of the HUT team's paper in PASP (<http://adsabs.harvard.edu/abs/2013PASP..125..431D>). New HUT spectral container files were created for our SSAP VO service, and a new interactive plotting tool was introduced to allow users to explore the spectra (zoom, change axis ranges, etc.)

SWIFT UVOT

Swift UVOT data is ingested monthly into MAST from HEASARC following suggestions from the HEASARC users group and the SWIFT UVOT Team. Database tables of observations and exposures are created with which users can search, display, and retrieve, images of interest.

A MAST web site was also created for SWIFT with links to various web sites providing Swift documentation. An ongoing effort is also underway to calculate footprints for each Swift UVOT exposure, and expand the available metadata for users following the current standards from the IVOA. The added information will be incorporated into a new database and search interface to be made available later in 2013.

Community interaction

Planning for the annual MAST survey and Users Group meeting is underway and is currently planned for fall of 2013.

MAST has collaborated with the AstroBetter site administrators and have posted a blog post about MAST's community contributed High-Level Science Products. The blog informs the reader about the products already archived and also invites new contributors.

MAST has a Facebook page and a twitter feed that are used to announce new products and features at MAST, supplementing newsletters and web page announcements.

Outreach

MAST and the HST Office of Public Outreach (OPO) continued the work required to make select HST Press Release image available via VO, collaborating to integrate AVM standard meta-data tags into the press-release images and to create a database and associated Virtual Observatory (VO) web services.

Other Major work efforts

MAST staff worked on many projects during the past 7 months that introduced new or enhanced capabilities or attributes. We describe a few highlights below.

MAST “portal”

The MAST Data Discovery Portal is a web application for exploring astronomical data from MAST and from the Virtual Observatory (VO) community. The application utilizes a variety of filter and visualization techniques to enable identifying data of interest from among a very wide variety of collections. The Data Discovery Portal infrastructure can be deployed in multiple forms to suit the specific user needs.

In June 2013 the Portal was deployed as the VAO Data Discovery Tool v1.5, marking the final deployment for the VAO before that project closes out in September 2014. Going forward, all new features will be exposed only through MAST web pages.

In September 2013, the Portal will replace the cross-mission search feature on the MAST home page. This replacement required integration with the MAST cross-mission database (CAOM) and a new data download/distribution capability. Another major new feature is the ability to cross-match any target list with either MAST observations or a wide variety of catalogs available from the Centre de Données astronomiques de Strasbourg (CDS).

In addition, the Portal infrastructure has been used to design a new search capability for the JWST Scientific Instrument Integration and Test Data (SID) archive. This technique, which uses the Portal’s filtering interfaces to drive database queries, is generic enough that it will be used in the future for a variety of mission-specific MAST searches.

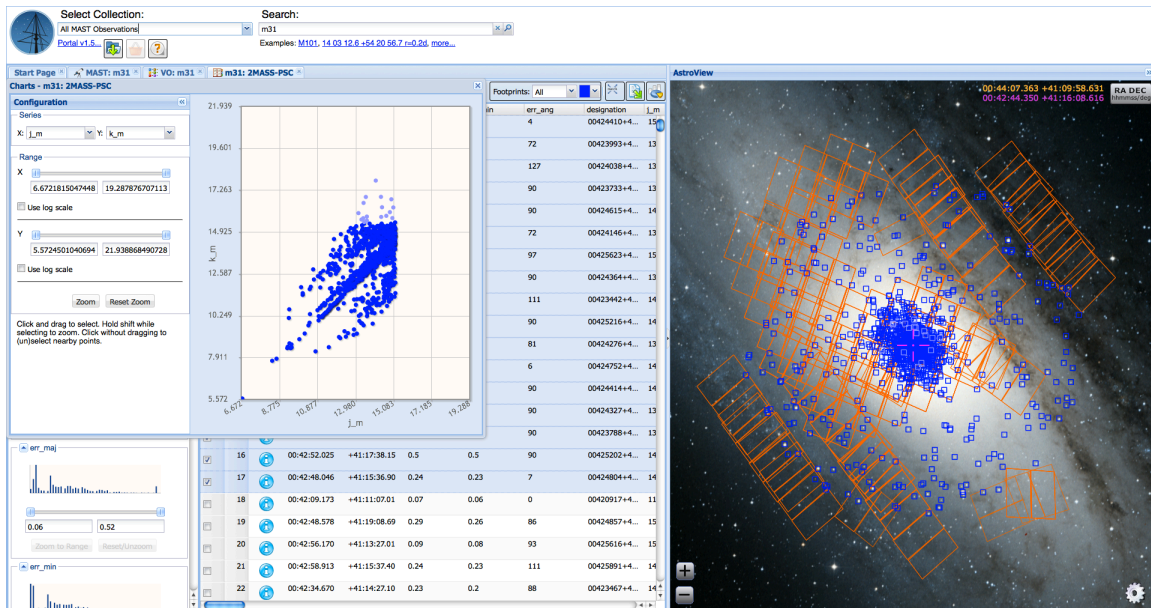


Figure 7: Snapshot of MAST Data Discovery Portal illustrating observational footprint and charting capabilities

Common Archive Observation Model (CAOM)

In collaboration with the Canadian Astronomical Data Center (CADC), we have been working to implement a common data model to describe available observational data. This enables users to search across multiple missions for related data in a uniform way and immediately provide results in a format compatible with the *Virtual Observatory* (VO), whilst reducing software development overhead. MAST is working on providing access to all of its holdings through this mechanism, as is CADC.

Hardware and Migration

At the end of 2012 MAST purchased a significant amount of SAN storage space – almost 500TB. Of this 250TB was to replace storage hardware that was end-of-life, 100TB was to accommodate planned increases in data holdings, and 150TB was to allow duplicate copies of operational databases in order to provide both backup copies and highly-available access services. We are currently migrating datafiles and databases to their new locations. As part of this process we are upgrading the network connections between servers and storage to 10GB/s to improve data access times.

We have now deployed operationally a cluster of SQLservers using ‘Always-On’ technology. This not only provides us a significant increase in capacity to handle demand but also provides 24x7 availability and data redundancy.

Community-Contributed High-Level Science Products (HLSP)

Twenty-four sets of community contributed reduced science ready data sets from 10 different projects were ingested into MAST during this reporting period.. Three of the projects are for Hubble Multi-Cycle Treasury Programs for which there have been multiple deliveries. Other teams that contributed data include the Hubble Heritage project data for the horsehead nebula and Comet ISON; the HST Orion Treasury project, the eXtreme Deep Field, IUE White Dwarf Spectral Atlas and additional deliveries from the ACS Globular Cluster and Brightest of Reionizing Galaxies (BoRG) teams MAST ingested 2.9 TB of HLSP during this time bringing the total HLSP archive to just over 4.95 TB. During the same time period, over 13.5 TB of HLSP were distributed to over 3600 users, demonstrating that these types of products remain very popular.

