

NASA Data Center Annual Program Plan
NASA Grant Number [NNX13AC07G](#) / STSci Grant J1561
September 2013 through August 2014
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, 2014 MAST holdings are over 300 TB, including 70 TB of Hubble Legacy Archive data, over 101 TB of HST standard pipeline products and over 6 TB of Community Contributed High Level Science Products. Figure 1 shows MAST holdings as of Sep 1, 2014.

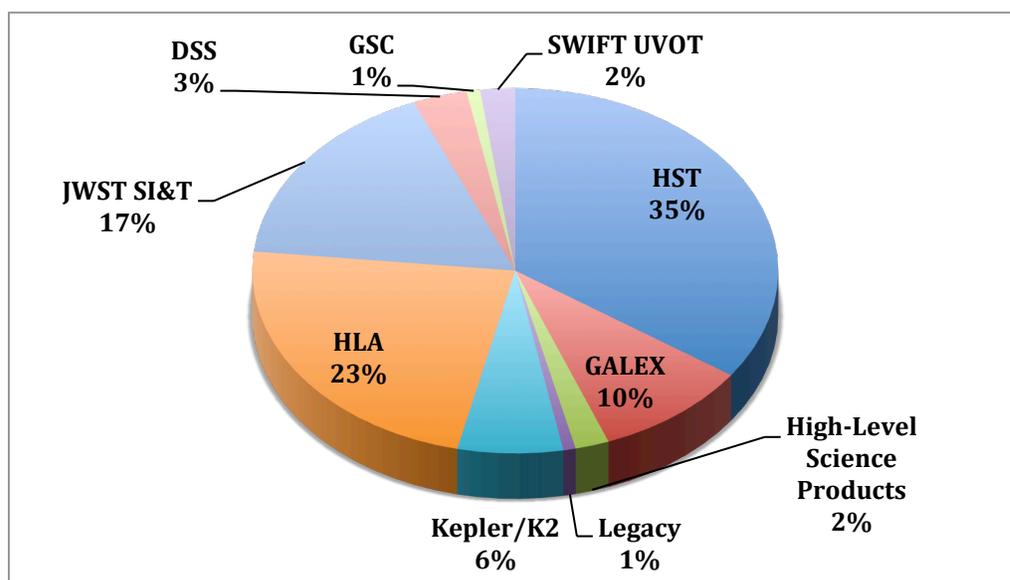


Figure 1: MAST Holdings

The archives at STScI have distributed over 385 TB of data between September 2013 and the end of August 2014, and have ingested over 144 TB of data. The figure below shows the statistics on data ingest and distribution to the public from Jan 1995 through August 2014.

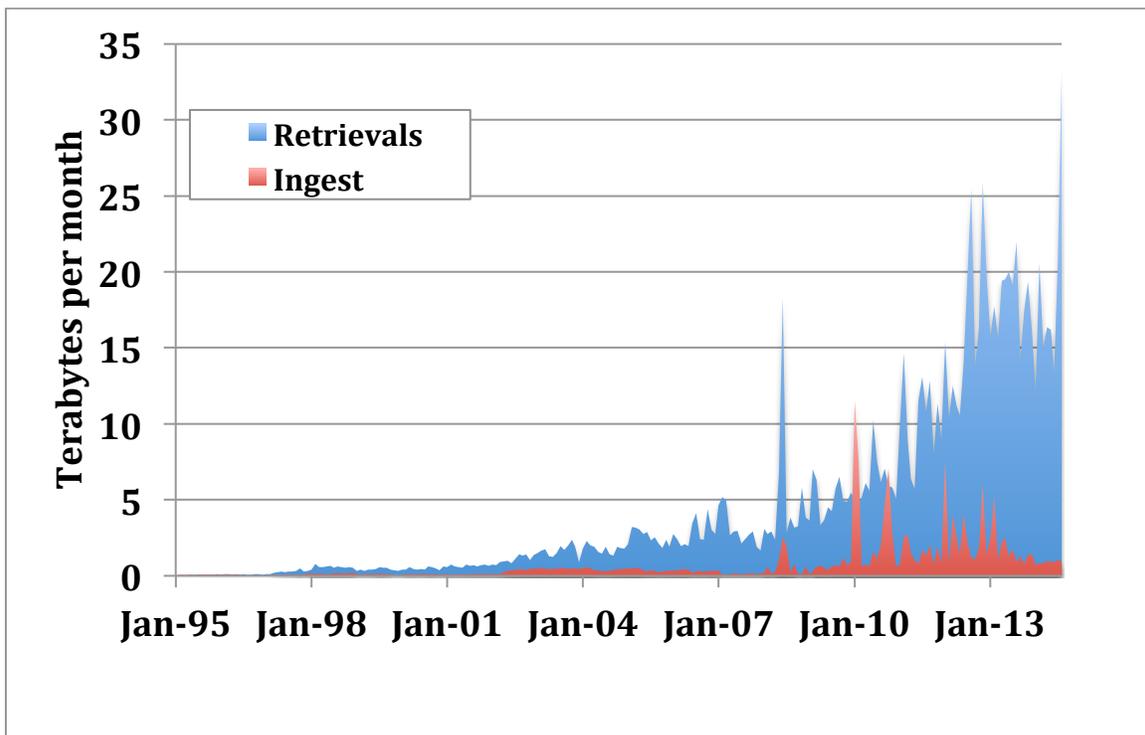


Figure 2: MAST Ingest and Distribution

Mission/Project reports

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

The Hubble Legacy Archive (HLA) project generates higher-level science products for the Hubble imaging instruments (WFPC2, ACS, WFC3, NICMOS). The data products include both improved images and source lists of objects in the images. The Hubble Source Catalog (HSC) project is taking the HLA source lists and integrating them into a combined catalog that matches measurements of objects taken at different times in different filters. The HSC summary catalog has one entry for each matched object with the object's mean position and average brightness in various filters; the HSC detailed

catalog has complete information on all measurements of that object's magnitudes, positions, and other information.

Most work on the HLA over the past year has been oriented toward providing better quality source lists for incorporation into the HSC. In October 2013, HLA Data Release 7.2 was made public. The main product in that release was improved source lists for the WFC3 data, including both better quality and removal of spurious sources through more sophisticated flagging. Since then the HLA has been working toward the next data release (expected in November 2014) which will include a complete reprocessing of WFC3 data with twice as many images and with much improved source lists. The new WFC3 pipeline is based on the latest version of the Drizzle software, Astrodrizzle, which will make it easier to support for all the HST images going forward.

The new DR7.2 source lists were incorporated into the Beta 0.3 version of the HSC, which was released in March 2014. That version of the HSC was of much improved quality compared with the Beta 0.2 version (May 2013). The HSC project is now working toward a version 1.0 release in early 2015 that incorporate the newest HLA source lists along with additional improvements in the catalog matching and construction.

Galaxy Explorer (GALEX)

Post-NASA AIS and GI Data

All of the post-NASA (CAUSE) raw data are at MAST, as well as all of the pipeline products produced by the mission (which amount to a subset of the raw data taken). In total, the CAUSE raw data amounts to ~3400 eclipses (~2 TB). MAST has already released the raw and pipeline products for all eclipses observed as part of Jamie Lloyd's Kepler GI CAUSE program. We plan to release the remaining CAUSE raw data, as well as all raw data from the pre-CAUSE (GR7) mission, in the next few months. These data will be useful for advanced users who want or need to perform their own calibrations starting from the raw data.

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.

Kepler Preview Plots

MAST has generated PNG preview plots for every Kepler lightcurve across all Quarters. These preview plots are designed in four different sizes (large, medium, small, and thumb), and are used prominently in the MAST Discovery Portal to allow users to get a quick visual preview of what a target's lightcurve looks like prior to downloading data.

K2 EPIC Catalog Released

The revitalized Kepler mission, now known as K2, began officially in June 2014. MAST has released the EPIC catalog of sources, from which a majority of targets observed by K2 are selected. The catalog (currently ~11 million rows) is available through a MAST Classic Search Interface as well as through the Kepler CasJobs SQL Interface.

K2 Campaign 0 Data Released

The first “campaign” (~75-day observing runs centered on a particular pointing along the ecliptic plane) is referred to as Campaign 0. The data, consisting of target-pixel files for nearly 8,000 targets that are similar in format to Kepler target-pixel files, were made public through MAST in Sept. 2014, via a Classic Search Interface.

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. The SWIFT data were added to the Common Archive Observation Model (CAOM) database, with the appropriate meta-data and foot print information. There is now a standard procedure for updating the data, footprints and databases.

Community interaction and outreach

Planning for the annual MAST survey and Users Group meeting is underway and is currently planned for December 2014.

The 2013 MAST Survey was run in October 2013. Results were summarized and discussed during the MAST Users Group (MUG) and are available at <http://archive.stsci.edu/surveyresults/>. The MAST Users Group met November 18-19. Information about the membership, and the MUG report are posted online at <http://archive.stsci.edu/mug/pastmug.html>.

The AVM Tagging Project went through a refactoring effort to streamline the image tagging and archival ingest workflow. This project is tasked imbedding 3-color composite press release images taken with the Hubble Space Telescope with metadata based on the AVM Standard (http://virtualastronomy.org/avm_metadata.php). This metadata contains descriptive information such as target category, Hubble instrument, filter, exposure date and position in the sky. Through a cooperative effort between STSci Office of Public Outreach (OPO) and the Mikulski Archive for Space Telescopes

(MAST) development team, over 1600 Hubble press release images were tagged and loaded into the MAST archive for permanent storage and retrieval. Much of the tagging is a manual effort requiring a trained specialist to derive the meta tags based on the initial science observations. After the new streamlined workflow was established, the OPO team performed the manual tagging effort and subsequent auto ingest into MAST. The images are now available through multiple web sites, including the MAST Web Portal (<http://mast.stsci.edu/stpr>) and the AstroPix public archive (<http://astropix.ipac.caltech.edu/link/1et>) and the World Wide Telescope (<http://worldwidetelescope.org/>).

Other Major work efforts

MAST Portal

The MAST Discovery Portal (mast.stsci.edu) was deployed in Fall 2013 both to replace the cross-mission search feature on the MAST home page, and to support Build 2 JWST Data Management System. This web application offers one-stop access to data from all of MAST's supported missions and from Virtual Observatory resources around the world. A sky viewer puts data in context by overlaying observation footprints and catalog objects on backgrounds from DSS and other surveys. A robust filtering mechanism allows data of interest to be identified and selected from the many search results.

Subsequent deployments in January and July 2014 solidified the infrastructure by improving performance, addressing user feedback, creating tutorial videos, upgrading support libraries, and enhancing usage logging. The July release also added access to more data sets such as Kepler FFI images, GALEX grism images and IUE high dispersion spectra.

With the VAO project ending in September 2014, MAST will become a member of the newly formed Virtual Observatory Alliance. As part of that collaboration, a new version of the VAO Data Discovery Tool was deployed in August 2014 to include many of the newer features of the MAST Discovery Portal.

Several projects are underway for release over the next year. These include Hubble Source Catalog search and display, interactive spectral plots, advanced observation search, column arithmetic, user login, and the addition of more background surveys to the sky viewer.

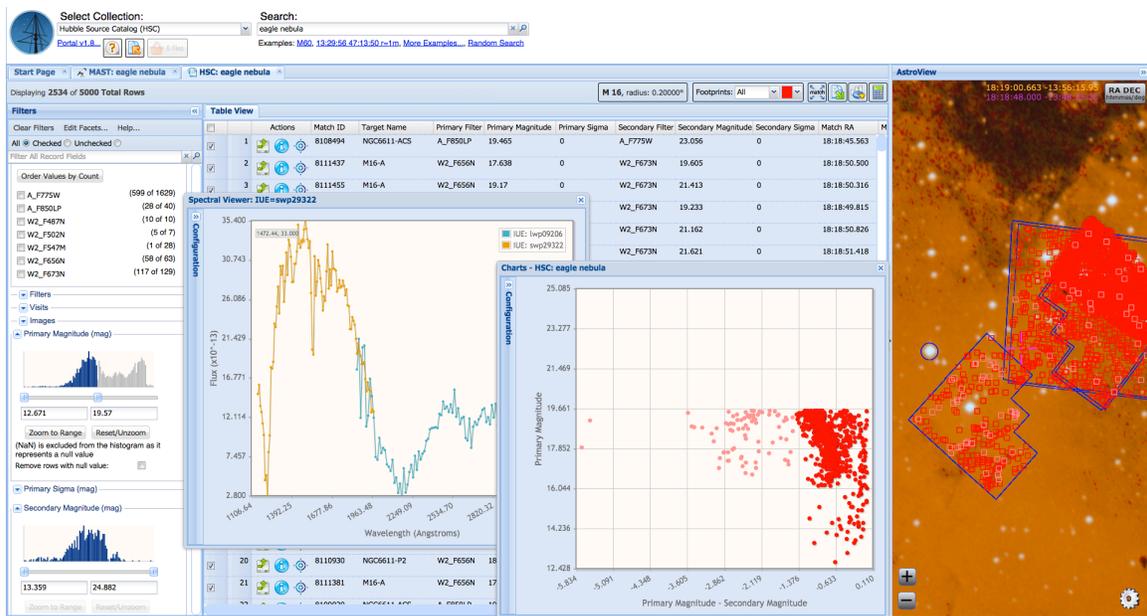


Figure 7: MAST Discovery Portal displaying Hubble Source Catalog results, filtered to show the brightest objects, plotted with a created column (Primary Magnitude - Secondary Magnitude), and including a plot of two IUE spectra.

Common Archive Observation Model (CAOM)

We are continuing our collaboration with the Canadian Astronomical Data Center (CADC) on a common data model to describe available observational data.

This has the benefit of enabling cross-mission data searches and makes it easy to return results in a VO-compatible format.

The initial prototype model has been populated with metadata from most of the MAST missions and is being used by the new MAST data discovery portal.

After some experience with this model, both CADC and STScI have recently adopted an improved version with additional metadata, and work is in progress to update our operational version.

Hardware and Migration

At the end of 2013 MAST purchased 150TB of SAN storage space to both replace some end-of-life hardware and increase capacity to store the GALEX photon files received at mission closeout.

MAST also retired and replaced several end-of-life servers as part of our transition to high-availability services and databases with hardware redundancy and automatic fail-over.

Vocabulary/AstroTag

AstroTag is a service in development by MAST staff to provide a browsable interface to archive data using a hierarchical concept tree or thesaurus. This service provides an alternative point of entry to retrieve archive data by category, rather than requiring the user to know and enter a specific object name, ID number or set of coordinates in order to access data.

In July 2014, MAST staff began a redesign of the AstroTag service, with the goal of improving the thesaurus currently used by AstroTag and the manner in which stored data is tagged in the database. In August 2014, MAST staff performed an evaluation of the Unified Astronomy Thesaurus (UAT), a community-supported, collaboratively edited thesaurus designed to incorporate many disparate thesauri across the astronomy community (<http://astrothesaurus.org>), as a possible replacement for the IVOAT thesaurus currently used.

AstroTag currently uses direct text matching to link concepts to metadata associated with documents (typically academic papers or HST proposals), along with a database of associations between documents and data sets in order to link concepts to data. To evaluate and compare IVOAT and UAT, we looked at how well the concept labels from each thesaurus could be used to retrieve documents via their metadata (titles, abstracts or keywords). We found that the IVOAT provides more coverage (about 100% of papers and 99% of HST proposals contain terms from the IVOAT, compared to 99% of papers and 87% of HST proposals for UAT), but the UAT is more user friendly. (E.g., the UAT contains 15 top level concepts, whereas the IVOAT has 274.) The output of the thesaurus analysis also identified hundreds of terms from both thesauri that never appear in any document and could likely be excluded from the final controlled vocabulary used for search in the system.

In the near future we plan to explore software solutions for expanding the UAT thesaurus with additional labels to boost coverage, by leveraging subject-matter expertise from research scientists on staff, as well as data from the thesaurus analysis already completed. For 2014-2015, the AstroTag redesign will include an interface for the Mast Data Discovery Portal.

Community-Contributed High-Level Science Products (HLSP)

Several new High-Level Science Products (HLSPs) were delivered and archived by MAST over the past year. We summarize these HLSPs below.

HST Mosaics of Comet ISON

Comet ISON made a splash in the general news circuit with its close approach to the Sun.



Figure 3: HST WFC3 two-color mosaic of Comet ISON from Nov. 2013.

Several HST campaigns were planned to study this comet as it passed through periastron, and calibrated mosaics were released through MAST as an HLSP. A total of four mosaics have been released, corresponding to observations taken in June, July, October, and November by HST using WFC3.

GCAT and BCSCAT GALEX Unique Source Catalogs

Both the GCAT and BCSCAT unique source catalogs provide “best” (deepest) fluxes for millions of GALEX sources by properly taking into account tile overlaps. The three GCAT catalogs of Seibert et al. use data from GR6 AIS, GR6 MIS, and GR7 Kepler area, respectively. The two BCSCAT catalogs of

Bianchi et al. use data from GR7 AIS and GR7 MIS, respectively. All catalogs are available through the MAST GALEX CasJobs SQL interface.

HST-COS Intergalactic Medium Spectral Survey

Danforth et al. (2014) used HST COS spectra of sightlines towards 75 bright AGN to study the intergalactic medium (IGM). After coadding COS spectra to obtain the highest signal-to-noise possible, they identified in a semi-automated fashion numerous absorption features coming from both galactic and intergalactic sources, and use the redshifted intergalactic features to identify intergalactic clouds and study the IGM. Both the coadded spectra as well as the linelists of absorption features were provided to MAST as an HLSP. MAST also provides interactive plots of the coadded spectra that allow users to zoom and overplot line identification markers of absorption features.

HLA Star Clusters

Compact star cluster catalogs of 20 nearby star-forming galaxies from Whitmore et al. were released as a MAST HLSP / HLA data product. Using HLA-generated catalogs from HST ACS observations, the team was able to study the luminosity function using star clusters selected in a more automated approach compared to previous studies. MAST provides access to both the HLA ACS images and the star cluster catalogs.

UV Ultra Deep Field (UVUDF)

Teplitz et al. (2013) used HST WFC3 to provide deep near-UV imaging of the Hubble Ultra Deep Field (UDF). Despite having deep observations in the far-UV, optical, and NIR wavelengths, the UDF was lacking in deep near-UV data. The UV-UDF program obtained point source detection limits down to a factor of 10 deeper than the previous best-available coverage (through GALEX observations). MAST provides access to the image mosaics, photometric catalogs, and improved redshifts based on these new, near-UV observations of the UDF as an HLSP.

GALEX all-sky maps

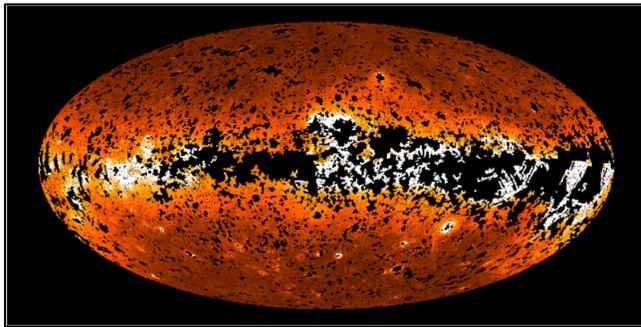


Figure 4: NUV diffuse radiation map from Murthy et al.

Foreground emission maps were provided by Jayant Murthy (Murthy, J 2014, *ApJS*, 213, 32). These all-sky maps provide measurements of the diffuse UV background, as well as geocoronal and zodiacal foreground emission, on 2-arcminute spatial scales, and are available as both a MAST HLSP page as well as the GALEX CasJobs SQL.

G191-B2B White Dwarf Spectral Line Atlas

Preval et al. (2013) have co-added FUSE and HST spectra to create one of the highest signal-to-noise spectra of a white dwarf to-date. Combined with the latest spectral linelist databases, the team was able to identify nearly 950 absorption features and measure abundances for nine different elements within G191-B2B's atmosphere. Their comprehensive white dwarf linelist will be valuable in abundance work on fainter white dwarf spectra. The coadded spectra, individual spectra, and linelist catalogs are provided as an HLSP.

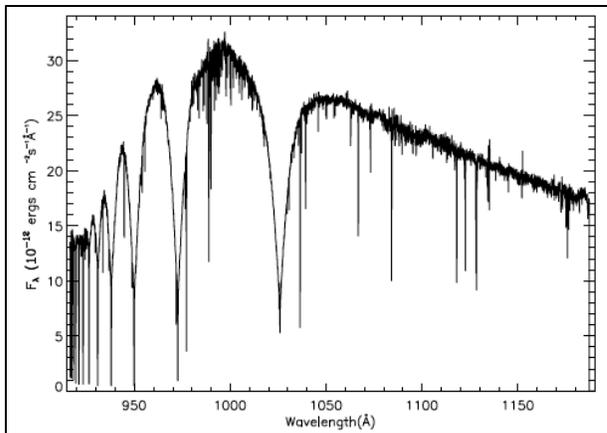
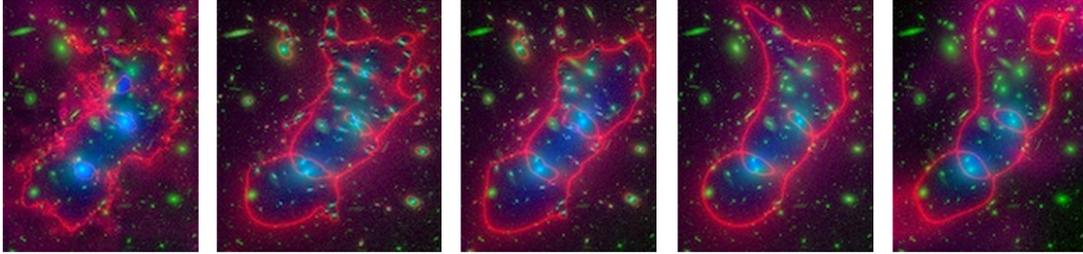


Figure 2: FUSE coadded spectrum of G191-B2B from Preval et al. (2013).

HST Frontier Fields

The Frontier Fields project is an HST initiative of six deep fields centered on strong lensing galaxy clusters in parallel with six deep "blank fields" adjacent to these clusters. The Frontier Fields project has released mosaics for Abell 2744 and MACSJ0416.1-2403. In addition, several lensing models have been released as HLSP. The archive is providing access to the data and providing interactive displays for these data.

Abell 2744: Overlay of magnification (red) and mass models (blue) on the full-band HST imaging (green)



Bradač et al.

CATS Team

Merten, Zitrin et al.

Sharon et al.

Williams et al.