

NASA Data Center Annual Program Plan
NASA Grant Number [NNX13AC07G](#) / STScI Grant J1561
September 2014 through August 2015
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, 2015 MAST holdings are over 500 TB, including 90 TB of Hubble Legacy Archive data, over 113 TB of HST standard pipeline products and over 8 TB of Community Contributed High Level Science Products. Figure 1 shows MAST holdings as of Sep 1, 2015.

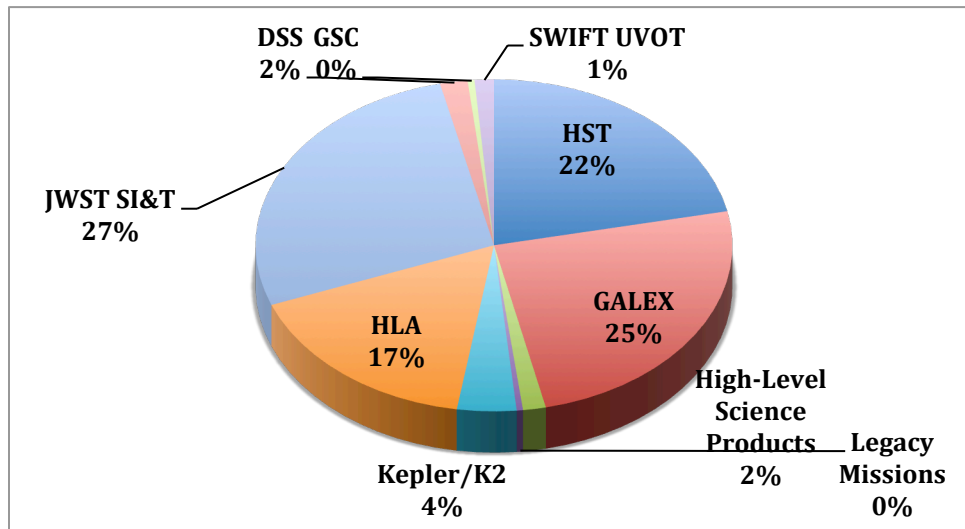


Figure 1: MAST Holdings

The archives at STScI have distributed nearly 254 TB of data between September 2014 and the end of August 2015, and have ingested nearly 74 TB of data. This ingest does not include the population of a 150 TB database of GALEX photons for the gPhoton project. An early release for gPhoton was in August 2015. The figure below shows the statistics on data ingest and distribution to the public from Jan 1995 through August 2015.

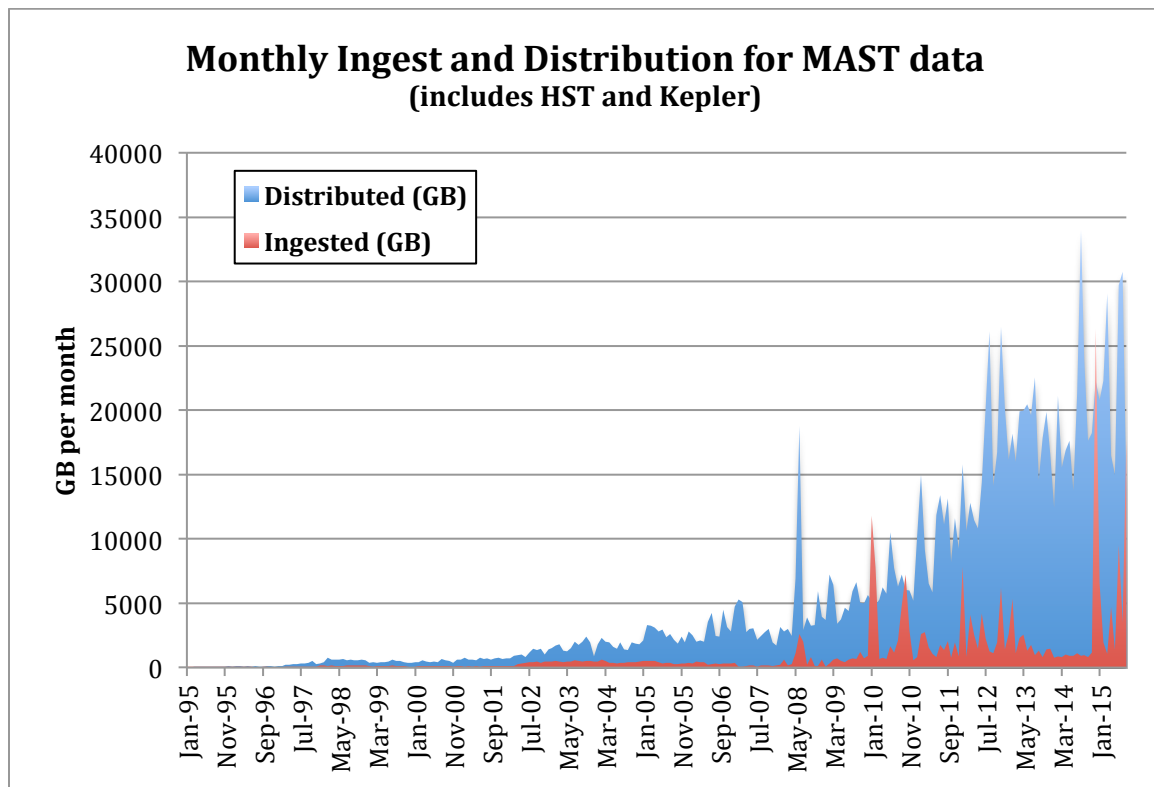


Figure 2: MAST Ingest and Distribution

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. On December 3, 2014 the HLA released DR8. In this release we improved WFC3 image data and source lists by using Drizzlepac and Astrodrizzle. This addition improves the image alignment and quality, thus providing increased number of datasets. It also allows source lists to go

deeper. There were user interface enhancements to this release as well. For visits that have more than 3 filters available, the color image entries have been modified so that the interactive display has access to all those images. New functionality to better support overlaying Hubble Source Catalog (HSC) was also in this release.

On February 25, 2015 the HLA released DR8.1. This release was entirely to support the HSC Version 1 release. A new higher performance approach is being used for the HSC overlay query in the interactive display. This supports catalogs with up to 50,000 sources with reasonably good performance. The HSC overlay catalog now includes all filters that have data and has some additional columns. The advanced HSC controls in the interactive display now include an option to choose whether MAGAUTO or MAGAPER2 magnitudes are shown.

The HLA is continuing to refactor the image and source list pipelines to include Drizzlepac and Astrodrizzle. In the fall of 2015 we will release DR9, which will be reprocessed and new ACS data using the new pipeline.

The ***Hubble Source Catalog (HSC)*** is designed to combine the tens of thousands of single visit-based source lists from the Hubble Legacy Archive (HLA) into a single master catalog. This provides entry of Hubble into the world of database astronomy, as pioneered by the Sloan Digital Sky Survey (SDSS).

The release of Version 1 of the HSC in February 2015 represents a major milestone for the project. The HSC currently includes WFPC2, ACS/WFC, WFC3/UVIS, and WFC3/IR photometric data generated using SExtractor software to produce the individual source lists. The catalog includes roughly 80 million detections of 30 million objects involving 112 different detector/filter combinations, and about 160 thousand HST exposures. Source lists from Data Release 8 of the Hubble Legacy Archive provide the underlying photometric and astrometric information. The primary method for accessing the HSC is the MAST Discovery Portal. However, for larger or more detailed queries a CasJobs (Catalog Archive Server Jobs System) interface.

The mean photometric accuracy for the catalog as a whole is better than 0.10 mag, with relative accuracy as good as 0.02 mag in certain circumstances (e.g., bright isolated stars). The relative astrometric residuals are typically within 10 mas, with a value for the mode (i.e., most common value) of 2.3 mas. The absolute astrometric accuracy is better than ~ 0.1 arcsec for most sources, but can be much larger for a fraction of fields that could not be matched to the PanSTARRS, SDSS, or 2MASS reference systems.

A paper has been submitted to the *Astronomical Journal* describing the database design with emphasis on those aspects that enable the users to fully exploit the catalog while avoiding common misunderstandings and potential pitfalls. Usage examples to illustrate some of the science capabilities and data quality characteristics are also included.

Galaxy Explorer (GALEX)

Initial Release of the gPhoton Database and Software Tools: The gPhoton database and software package has been released for public use. The database contains every photon event observed by the spacecraft through GR6/7, amounting to over a trillion rows (130 TB on disk). Open-source python software is provided to allow anyone to query this database and construct calibrated lightcurves and images with user-specified apertures and time bins. In particular, this enables studies of inter-visit variability at the timescales of seconds and minutes, and also simplifies construction of lightcurves and images for fields-of-view that include overlapping GALEX visits. In addition to improving query efficiency, backend reliability, and astrometric/photometric calibration, we will also be adding new features and capabilities to the software over the coming year. A paper describing the project is in advanced stages and expected to be submitted in Fall 2015.

Kepler/K2

The Kepler Data Management Center (DMC), funded directly from the Kepler Project, was established at STScI to archive the Kepler Data products. The past year has seen data releases from K2 Campaigns 0 through 4. 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.

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

The annual MAST Survey was administered during November 2014. 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 December 2-3, 2014. Information about the membership, and the MUG report are posted online at <http://archive.stsci.edu/mug/pastmug.html>.

Planning for the next annual MAST survey and Users Group meeting is beginning. The MAST Users Group (MUG) meeting is currently planned for January 2016, to better fit the MUG member's schedules. This year the new chair is Sarbani Basu.

After the new streamlined workflow for the AVM Tagging Project was established last year, the OPO team performed the manual tagging effort and subsequent auto ingest into MAST. There are now 1700 images available through multiple web sites, including the MAST Web Portal (<http://mast.stsci.edu/stpr>), 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

During the second year of operation of the MAST Discovery Portal (<https://mast.stsci.edu>), 3 new major versions and 13 minor versions were delivered to provide a variety of enhancements and fixes.

In February, 2015, version 1.8 included several features to support searching and retrieving data from the Hubble Source Catalog (HSC). These features included high resolution HST image overlays in the AstroView sky viewer, cutout preview images around sources of interest, data column arithmetic to support plotting, and the ability to crossmatch against MAST catalogs (HSC, GALEX, SDSS and 2MASS so far). This version also included interactive spectral plots, and major enhancements to logging and integration with the log analysis tool Splunk (<http://www.splunk.com>).

In April, 2015, version 1.9 was delivered to support Build 4 of the JWST Data Management System (DMS). In this version, the Portal allowed user login by integrating with the STScI Shibboleth Single Sign-On infrastructure.

In July 2015, version 2.0 included a significantly retooled user interface. These interface changes were driven by a usage study conducted in cooperation with the STScI Office of Public Outreach (OPO). Additional new features were time series and light curve viewers, and multiple new background image sets for AstroView such as Wise, Plank, Fermi, Spitzer and WMAP.

Work was also done this year on an advanced observation search interface with release expected this fall. Other major projects slated for release in the coming year are advanced catalog search, user login to support display and retrieval of proprietary data, support for JWST DMS builds 5 and 6, and consolidation of some MAST web pages by retiring the GalexView and HLA search interfaces in favor of the Discovery Portal.

NASA Virtual Observatory Project

MAST continues to participate the NASA VO project (NAVO) and the International Virtual Observatory Alliance (IVOA). Theresa Dower is now the deputy chair of the IVOA Registry working group and Tom Donaldson is the deputy chair of the IVOA Data Access Layer Working Group. The IVOA registry is located and maintained at STScI. MAST staff members are collaborating with IPAC staff to compare and report on multiple spatial indexing schemes for point-source and extended-object datasets. MAST continues to support a variety of VO services and plans to review them over the next year to ensure they are complete and accurate.

Common Archive Observation Model (CAOM)

We are continuing our collaboration with the Canadian Astronomical Data Center (CADC) and the European Space Astronomy Center (ESAC) on a common data model to describe available observational data. We have recently started discussions with IPAC on sharing observational metadata using this same model. This has the benefit of enabling cross-mission data searches and makes it easy to return results in a VO-compatible format. The MAST discovery portal now uses this model to search observations from all MAST missions. There is ongoing work to streamline the metadata collection into a pipeline process that can provide automated inclusion of new datasets into the databases.

Hardware and Migration

MAST retired and replaced several end-of-life servers as part of our adoption of high-availability services and databases using hardware redundancy and automatic fail-over. MAST also purchased 250TB of storage space primarily to replace some end-of-life disk hardware. A number of network upgrades to improve connectivity to network storage was also completed

Vocabulary/AstroTag

AstroTag is a service in development by MAST staff to provide a browsable interface to archived 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 the past year, the AstroTag effort has focused on improving the AstroTag tag set, as well as creating a data model for AstroTag annotations.

The tags that will be used for AstroTag come from the Unified Astronomy Thesaurus (UAT). The UAT is a community-supported, collaboratively edited thesaurus meant to replace older astronomy thesauri and journal keyword sets currently used in the astronomy community. In February 2015, MAST staff participated in the Workshop on Shared Curation of Astronomical Literature, a workshop held at the Harvard Center for Astrophysics (CfA) library that in part focused on the UAT. During this meeting, MAST staff helped to plan a UAT clean-up project that is currently underway and will continue

through early 2016, in anticipation of the first official release of the UAT. The multi-institution clean-up project pairs individuals with backgrounds in library science and taxonomy with subject matter experts to collaboratively edit and improve sections of the thesaurus. MAST staff have already run several editing sessions with STScI research scientists. MAST staff members also helped to create a best practices document for thesaurus editors on the project, and are working closely with the STScI library, Harvard CfA library and AAS Journals to coordinate the project.

The data model in development for AstroTag is based in part on the draft Web Open Annotations data model currently under standardization by the W3C Web Annotation Working Group. The data model is designed to be expansible, so that additional collections of annotations could be developed in addition to semantic tags from the UAT. The model also includes the notion of provenance, requiring each annotation to have a source. For example, an annotation that is derived from keywords appearing in an associated paper would be annotated with an id that identifies the paper (or papers), so that this data can be displayed to the user for additional context. Along with the data model, a prototype web interface was developed that allows users to browse tagged data by data set, tag, or annotation source.

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.

K2 Extracted Lightcurves (“K2SFF”)

The lightcurves from K2 contain larger systematics than the original Kepler mission, due to the reduction in pointing precision as a result of having to rely on only two reaction wheels. Vanderburg & Johnson have created a technique to correct for the pointing-dependent nature of the pixel-level fluxes. The team released their extracted lightcurves for the first three K2 campaigns and also for the engineering campaign.

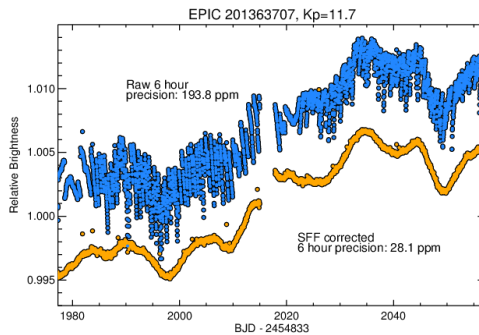


Figure 3 An example of the uncorrected fluxes from K2 (blue) and the corrected K2SFF version (orange)

K2 Variability Catalog (“K2VARCAT”)

Armstrong et al. have also created a technique to correct for the pointing-dependent nature of the pixel-level fluxes. In separate releases for campaigns 1 and 2 the team has released their extracted lightcurves (using a photometric aperture defined by the brightness of the target), as well as three plots of the resulting lightcurves: extracted, detrended, and phase-folded. The lightcurves are stored in FITS files a binary tables in

the first extension. The header also includes information on the aperture size used and the coordinates of the aperture center. The team has also released a catalog of variability statistics and object classification.

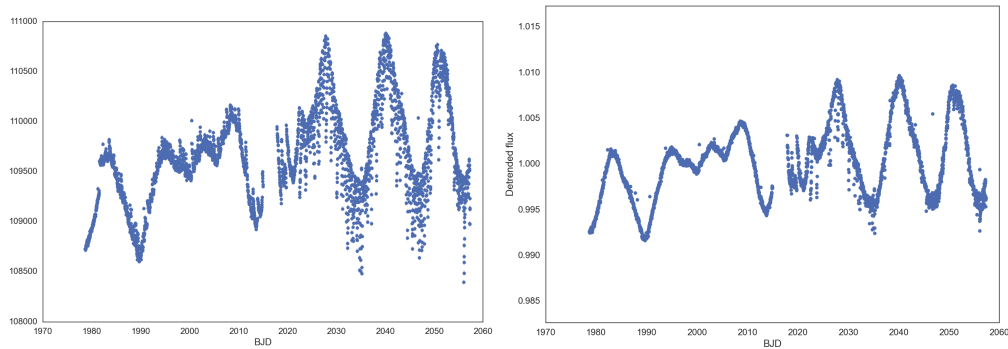


Figure 4: Extracted Flux and Detrended Flux

HST Frontier Fields

The Frontier Fields project is an HST initiative covering six deep fields centered on strong lensing galaxy clusters in parallel with six deep "blank fields" adjacent to these clusters. The project continued to release data for additional fields throughout the past year including data for MACSJ0416.1-2403, MACSJ0717.5+3745, and MACSJ1149.5+2223. New lensing models were also archived this year.

Cluster Lensing And Supernova survey with Hubble (CLASH)

The CLASH team archived several supplementary Subaru images and catalogs including: MACS 0744, MACS 0647, Abell 611, MACSJ1115+01, MACSJ1720+35, RXJ1532.9+3021, RXJ 2129, Abell 209, MACS 0429, MACS 1931, MACS 1311, MACS 1423, and MACS 2129.

Multi-Wavelength Imaging of the Eagle Nebula M16

One of the most iconic images from the Hubble Space Telescope has been the 1995 WFPC2 image of the Eagle Nebula (M16, sometimes known as the "Pillars of Creation"). Nineteen years after those original observations, new images have been obtained with HST's current instrumentation: a small mosaic in visible-light, narrow-band filters with WFC3/UVIS and a second infrared mosaic in broad-band filters with WFC3/IR. The Heritage team has provided the image mosaics as an HLSP.



Deep HST ACS Imaging of 47 Tuc and the SMC

Deep HST ACS images in two filters (f606w and f814w) targeting the globular cluster 47 Tuc have been released by Kalirai et al. as a new HLSP. These deep images go down to 30th magnitude, extending down to the faintest part of the main sequence and the coolest white dwarfs. In addition to 47 Tuc, there is overlap with the SMC, able to resolve stars down to 0.2 solar masses. Stacked images and source catalogs, as well as artificial source catalogs used for completeness tests, are also available.

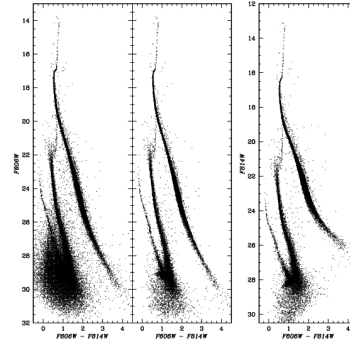


Figure 5: Color magnitude diagrams showing giants, dwarfs, and white dwarfs from 47 Tuc

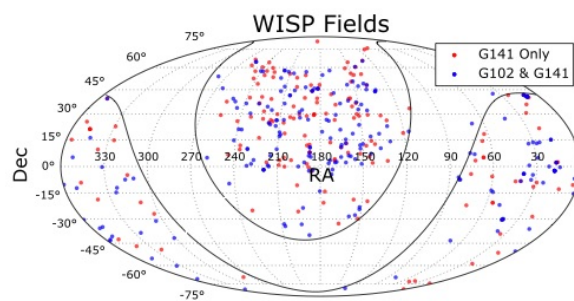
Ultraviolet UDF (UVUDF) Version 2 release

The Ultraviolet UDF team (PI: Teplitz; program ID 12534) has released v2.0 mosaics in F225W, F275W and F336W of the data obtained in epoch 3. In this release, all the UVIS data have been fully recalibrated and re-reduced using a combination of standard and custom calibration scripts and reference files, to improve the low-level noise structure in the final images. In addition to the image mosaics, an aperture matched PSF corrected photometric catalog is also made available, including photometric and spectroscopic redshifts in this field.

WFC3 Infrared Spectroscopic Parallel Survey (WISP)

WISP is an HST parallel survey using the WFC3 to collect images, grism spectra, and generate source catalogs in nearby areas of the sky while other HST instruments are in use. The primary science drivers include measuring the star formation rate of galaxies across 10

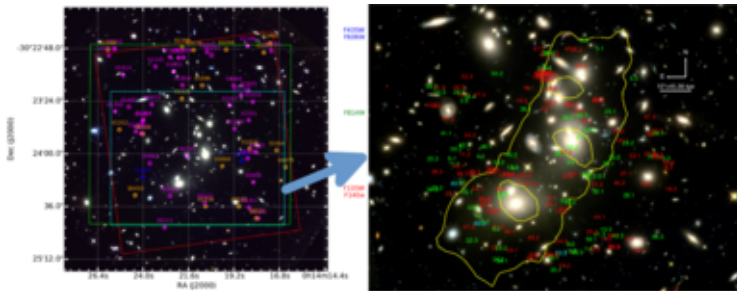
billion years of cosmic time, studying galaxy clustering on Mpc scales between redshifts $1 < z < 2$, constraining dust extinction/metallicity, and searching for $z > 6$ Lyman alpha emitters. The WISP team has released the images, extracted grism spectra and source catalogs for the first 136 parallel fields.



Grism Lens-Amplified Survey from Space ("GLASS")

GLASS is a Cycle 21 Large Program, consisting of 140 primary and 140 parallel orbits, targeting 10 massive galaxy clusters (including the 6 Frontier Fields clusters) using HST WFC3 and ACS grisms. The primary goal of the project is to obtain spectra

of faint galaxies with unprecedented sensitivity and angular resolution. Data of the first galaxy cluster has now been released at MAST (MACS 0717.5+3745). The data at MAST include processed and extracted spectra of targets in the field, source catalogs, redshift measurements, and LOTS of diagnostic plots.



LAPLACE

Schneider et al. archived a full recalibration of NICMOS coronagraphic images through Cycle 15. Some of these images have already been used to recover science previously unable to be realized, such as direct imaging of exoplanets or studies of circumstellar debris discs. Also included is the IDL-based "Image Display Paradigm 3" (IDP3) software suite for visualizing the data.

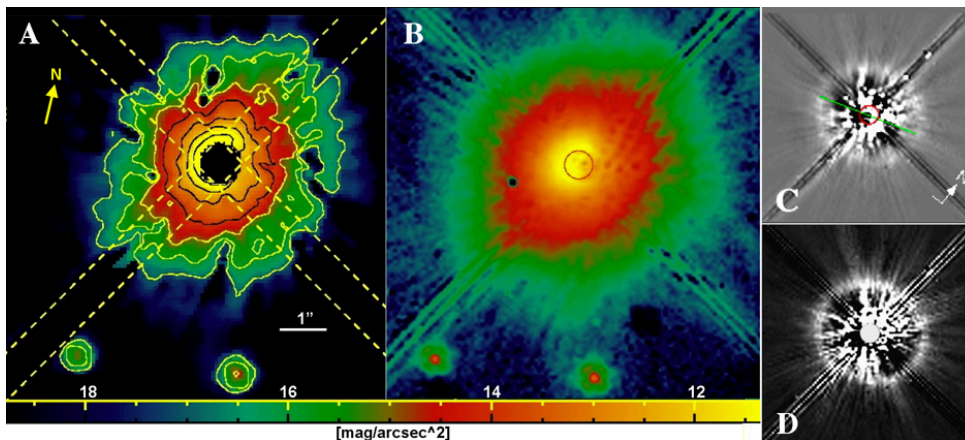


Figure 6: Representative improvements in CPSFTS imaging of circumstellar disks. (A) HD 100546 (F160W) as reported [Augereau et al. 2001](#), (B) with LAPL reprocessing and PSF templates. (C) HD 181327 (F110W) as reported by [Schneider et al. 2005](#), (D) with LAPL reprocessing and PSF templates.

M83 Mosaics

The M83 mosaic project (led by William Blair, JHU) produced beautiful mosaic images spanning seven fields across M83, using 15 filters from WFC3-UVIS and WFC3-IR.

Figure 8: A combined mosaic using the F555W and F547M observations

