



**MAST**

**Users  
Group  
Meeting**

**January  
23–24,  
2018**

# High Level Science Products: Current Status and Future Plans

Scott W. Fleming  
On behalf of the MAST team

Archive Sciences Branch  
Operations and Engineering Division





# MAST Users Group Meeting

January 23-24, 2018



**GLASS**  
Cycle 21 HST program to survey 10 galaxy clusters using grism spectroscopy. A total of 140 primary orbits and 140 parallel orbits.

**Modes**  
Lensing models from 7 teams for the six Frontier Fields. Includes shear, mass surface density, lensing deflection and image magnification.

**LEGUS**  
HST Treasury Program to image 50 nearby galaxies in multiple colors with ACS and WFC3. The galaxies are resolved into stars, clusters, and associations.  
Legacy ExtraGalactic UV Survey

**PHAT**  
A multi-cycle HST program to map 1/3 of Andromeda in 8 filters from the UV to NIR. The disk is resolved into > 100 million stars. The observations are grouped into 23 "bricks".

**UV-E**  
An all-sky mapping of diffuse UV background in FUV and NUV, GALEX data. Includes estimates of galactic and zodiacal foreground per pixel. Maps in arcminute grids.

**WISPR**  
A large HST pure-parallel program using WFC3 grisms to obtain images, spectra, and catalogs of star forming galaxies at redshifts  $0.5 < z < 2.5$ . A total of 385 fields will be observed.

**CLASH**  
An HST survey of 25 massive galaxy clusters using WFC3 and ACS. Includes mosaics, source catalogs, and lensing models, along with Subaru mosaics and

**Vanderburg et al. (K2SFF) and Armstrong et al. (K2VARCAT).**  
K2VARCAT also includes variability classification.

**MAST ARCHIVE @ SPACE TELESCOPES**  
archive.stsci.edu/prepds/

**SDSS**  
WFC3 Parallel

- Community-contributed data products
- Complement or supplement a MAST mission
- Any type of file (images, spectra, time series, models, catalogs, maps, linelists)



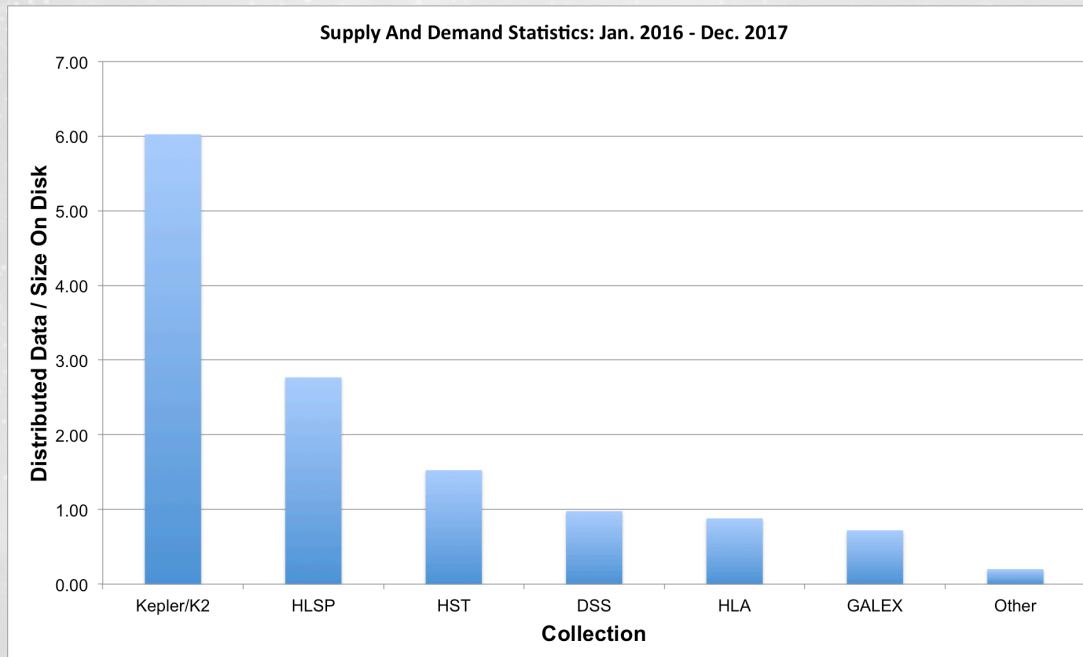


**MAST**  
**Users**  
**Group**  
**Meeting**

**January**  
**23–24,**  
**2018**

# Current Holdings (1/2)

- As a collection, HLSP are comparable to MAST's biggest missions in terms of community use
- Enables unique science from other MAST missions



- By design, new HLSP ingest is growing
- Growing number of HLSP who re-deliver new versions of existing HLSP
- Already close to limit we can support with the old ways we ingested HLSP





**MAST**  
**Users**  
**Group**  
**Meeting**

**January**  
**23-24,**  
**2018**

# Current Holdings (2/2)

Select a collection... and enter target:  
MAST Observations by Object Name or RA/Dec  Search

[About Collections...](#) [Show Examples...](#) [Random Search](#) [Advanced Search](#)

Upload Target List My Download Basket: 0 files [User Manual/Help](#) | [Leave Feedback](#) | [About This Site](#)

anonymous  
Login...  
Account Info...

Home Page **MAST: k2 201121245 r=0**

Displaying 6 of 15 Total Rows of Observations EPIC 201121245, radius: 0.00000°

**Filters** [List View](#) [Album View](#)

Clear Filters Edit Filters... Help... Edit Columns... Table Display: All Show Preview:  Show Cutout:

	Actions	Preview	Mission	Instrument	Project
<input type="checkbox"/>			K2	Kepler	hlsp_everest
<input type="checkbox"/>			K2	Kepler	hlsp_k2gap
<input type="checkbox"/>			K2	Kepler	hlsp_k2sff
<input type="checkbox"/>			K2	Kepler	hlsp_k2varcat
<input type="checkbox"/>			K2	Kepler	K2
<input type="checkbox"/>			K2	Kepler	hlsp_polar

**Dec (deg)**  
-06:08:21.64 -05:22:16.47

**Start Time (d)**  
2005-04-11 08:52:52 2014-07-22 14:38:59

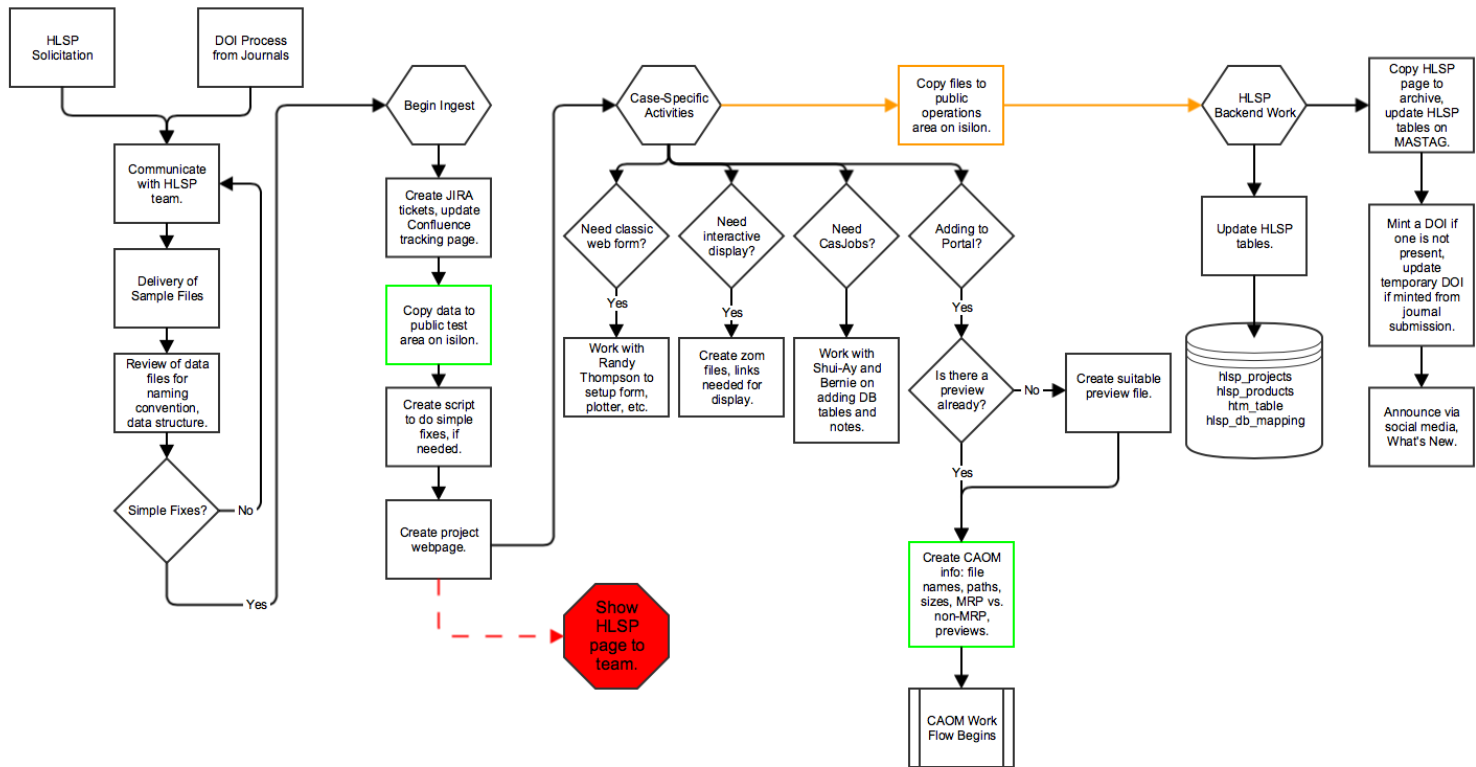
**End Time (d)**  
2007-03-20 10:54:50 2014-08-20 20:04:55

**Exposure Length (s)**

**AstroView** 11:48:02.756 -06:09:50.13 RA DEC  
11:47:18.743 -06:00:13.34 hhmms/deg



# Current Work: Improved Ingest



- Current HLSP process is manual, customized for each HLSP: **NOT SCALABLE**
- Some customization always needed (HLSPs are diverse by nature)
- Identified time-consuming parts of the process that can be standardized or automated with software development (minimize one-off scripts, web pages)
- See also improvements from other work we are doing relating to HLSP





**MAST**  
Users  
Group  
Meeting

January  
23-24,  
2018

# Current Work: Improved Ingest

openSAIL / MAST-HLSP

Unwatch 1 Star 1 Fork 1

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

Software for contributing High Level Science Products to MAST. **HLSP Guideline Compliance**

94 commits 1 branch 0 releases 2 contributors MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

scfleming - GUI has basic functionality, but needs major code cleanup. Latest commit 9e381ad 25 days ago

- CHECK\_DATA\_FORMAT - GUI has basic functionality, but needs major code cleanup. 25 days ago
- CHECK\_FILE\_NAMES - Fixes #3, allowing for files that are gzipped. a month ago
- PREP\_CAOM Single dataProductType now applied to all products. Removed from exte... a month ago
- docs - Renaming doc/ to docs/ 3 months ago
- LICENSE Initial commit 4 months ago
- README.md - Adding link to component documentation to primary README file. 3 months ago

README.md

## MAST-HLSP

This repository contains Python software used for contributing High Level Science Products (HLSP) to MAST. Tasks are sorted into different subdirectories within the repository, e.g., file name checking is located in one folder, keyword / data format checking in another, etc.

Task	Doc Link
check_file_names	<a href="#">DOC</a>

**Getting Stuff Into CAOM/Portal**





# Current Work: CAOM

- Most HLSP are *not* in CAOM/Portal.
  - Exceptions are K2 HLSP and some available through HLA.
- Working towards future where new HLSP go straight into CAOM/Portal on ingest, using Python code presented earlier.
- Also need to back-fill older HLSP into CAOM.
  1. Need to move old HLSP from current disk location to new location. Standardize file names, update DB tables, links+web pages, services.
  2. Need to create and test s/w to populate CAOM from diverse collection of data files. This is tied to development of CAOM v2.3.
    1. Phase 1 = Get Existing K2 HLSP into CAOM v2.3 using s/w.
    2. Phase 2 = ~6 diverse HLSP already available through HLA.
    3. Phase 3 = Rest of HLSP already available through HLA.
    4. Phase 4 = Rest of back-fill.

All the while, we will be developing our software to account for challenges due to HLSP diversity.





**MAST**  
**Users**  
**Group**  
**Meeting**  
**January**  
**23-24,**  
**2018**

# Current Work: DOI

## High Level Science Products

High-Level Science Products (HLSP) are community contributed, fully processed (reduced, co-added, cosmic-ray cleaned etc.) images and spectra that are ready for scientific analysis. HLSP also include files such as object catalogs, spectral atlases, and README files describing a given set of data.

Search below to find HLSP of interest by product, object type, and/or wavelength. **Select more than one item in each list by using the shift and the control keys.** Click on the "search" button for a list of the products for that project. The title of the project is a link to more information about the project. You may search for specific targets by using the [HLSP search page](#). You may also be interested in more [information about download options](#). MAST encourages the submission of HLSP based on data from its missions. Please consult the [Guidelines for Contributing HLSP](#) for more information.

Select Product Type	Select wavelength	Select Object Type	Select Mission
<input type="checkbox"/> Catalog <input type="checkbox"/> Composite <input type="checkbox"/> Image Atlas <input type="checkbox"/> Individual Object <input type="checkbox"/> Model <input type="checkbox"/> Spectral Atlas <input type="checkbox"/> Spectral Linelist <input type="checkbox"/> Survey <input type="checkbox"/> Time Series	<input type="checkbox"/> IR <input type="checkbox"/> Near IR <input type="checkbox"/> Optical <input type="checkbox"/> Ultraviolet <input type="checkbox"/> X-Ray <input type="checkbox"/> None	<input type="checkbox"/> AGN <input type="checkbox"/> Dust <input type="checkbox"/> Extragalactic <input type="checkbox"/> Galactic <input type="checkbox"/> Galaxy: Center <input type="checkbox"/> Galaxy: Clusters <input type="checkbox"/> Galaxy: Evolution <input type="checkbox"/> Galaxy: Field <input type="checkbox"/> Galaxy: Formation	<input type="checkbox"/> HST <input type="checkbox"/> FUSE <input type="checkbox"/> GALEX <input type="checkbox"/> K2 <input type="checkbox"/> IUE <input type="checkbox"/> HUT <input type="checkbox"/> SWIFT <input type="checkbox"/> KEPLER <input type="checkbox"/> TESS

[Catalogs](#) | [Composites](#) | [Image Atlases](#) | [Individual Objects](#) | [Models](#) | [Spectral Atlases](#) | [Spectral Linelists](#) | [Surveys](#) | [Time Series](#)

### Catalogs

- [30DOR: Hubble Tarantula Treasury Project](#) PI: Elena Sabbi [DOI](#)
- [3D-HST: Spectroscopic Galaxy Evolution Survey with HST](#) PI: Pieter van Dokkum [DOI](#)
- [ACSGGCT: The ACS Survey of Galactic Globular Clusters](#) PI: Ata Sarajedini [DOI](#)
- [ANDROMEDA: Deep Optical Photometry of Six Fields in the Andromeda Galaxy](#) PI: Thomas Brown [DOI](#)
- [ANGRRR: Archive of Nearby Galaxies: Reduce, Reuse, Recycle](#) PI: Julianne Dalcanton [DOI](#)
- [ANGST: ACS Nearby Galaxy Survey](#) PI: J. Dalcanton [DOI](#)
- [BCSCAT: Bianchi, Conti, Shiao \(BCS\) Catalog of Unique GALEX Sources](#) PI: Luciana Bianchi [DOI](#)
- [BIANCHI\\_GR5XDR7: Catalogs of GALEX UV unique sources, and of UV-optical matched sources](#) PI: Luciana Bianchi [DOI](#)
- [CANDELS: The Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey](#) PI: Faber/Ferguson [DOI](#)
- [CVARO: A Cataclysmic Variables and Related Objects Ultraviolet Spectral Catalog](#) PI: Patrick Godon [DOI](#)
- [CVCAT: A Catalog and Atlas of Cataclysmic Variables](#) PI: Ronald Downes [DOI](#)
- [DEEP47TUC: Deep HST ACS Imaging of 47 Tuc and the SMC](#) PI: Jason Kalirai [DOI](#)
- [DROPOUTS: Catalogs of B-, V-, and I-band dropout sources](#) PI: S. Beckwith [DOI](#)
- [GCAT: GALEX Unique Source Catalogs](#) PI: Mark Seibert [DOI](#)
- [GSWLC: GALEX-SDSS-WISE Legacy Catalog](#) PI: Samir Salim [DOI](#)
- [HLASTARCLUSTERS: The HLA Star Cluster Project](#) PI: Bradley Whitmore [DOI](#)
- [HPSGC: HST/NICMOS Paschen-alpha Survey of the Galactic Center](#) PI: Q.D. Wang [DOI](#)
- [KPLRUBV: Howell-Everett UVB Photometric Survey of the Kepler Field](#) PI: Howell [DOI](#)
- [PHAT: The Panchromatic Hubble Andromeda Treasury](#) PI: Julianne Dalcanton [DOI](#)

## Information about DOI 10.17909/T9RP4V

**DOI Creator(s):** Sabbi, Elena  
**Date:** 2016-01-22 05:00:00 Z  
**Title:** Hubble Tarantula Treasury Project  
**Display Data:** View data for [doi:10.17909/T9RP4V](https://doi.org/10.17909/T9RP4V)

**About this data:** HTTP is a panchromatic imaging survey of stellar populations in the Tarantula Nebula in the Large Magellanic Cloud that reaches into the sub-solar mass regime (<0.5 M<sub>☉</sub>). HTTP utilizes the capability of the Hubble Space Telescope to operate the Advanced Camera for Surveys and the Wide Field Camera 3 in parallel to study this remarkable region in the near-ultraviolet, optical, and near-infrared spectral regions, including narrow-band Ha images.

### Last Updated:

This DOI is provided by the MAST archive at Space Telescope Science Institute

- [DOI FAQs](#)
- [DOI Main page](#)

- All previous HLSP now have DOIs at the project level.
- All new HLSP will get a DOI minted as part of ingest.
- Working with journals on making sure these DOIs get inserted into papers by authors (not just for creation of new HLSP either).



**MAST**  
**Users**  
**Group**  
**Meeting**

**January**  
**23–24,**  
**2018**

# Current Work: Journals

Referral question from EJPress submission form (to follow existing EJPress question about data analyzed):

Do you wish to contribute data created and/or re-processed for this paper as a MAST High Level Science Product (HLSP)? (see mouse over text below)

If you are interested in this opportunity or need more information, fill in this brief form (link to webform)

[Suggested Mouse-over text](#) (up to EJ Press if they wish to integrate as mouse over text for "HLSP" in initial question, or leave as a brief paragraph here):

HLSPs are observations, catalogs, or models that complement, or are derived from, MAST-supported missions. These include Hubble, James Webb, Kepler/K2, GALEX, TESS, Swift, XMM, and others. HLSPs can include images, spectra, light curves, maps, source catalogs, or simulations. They can include observations from other telescopes, or MAST data that have been processed in a way that differs from what's available in the archive. HLSPs are permanently archived at MAST, get their own project webpage, and will appear in MAST search interfaces along with bibliographic references to your paper(s).



## Information about DOI 10.17909/T9010Z

**DOI Creator(s):** Pirzkal, Norbet  
Malhotra, Sangeeta

**Date:** 2017-08-01 04:00:00 Z

**Title:** Faint Infrared Grism Survey (FIGS)

**Display Data:** *Data products are not yet available. Once the data products are delivered and ingested into MAST, the final data products will be linked to here.*

**About this data:** The reduced and calibrated FIGS spectra will be made available via the MAST data archive and will include the single-PA 2D data stamps, the single PA extracted spectra using optimal extraction, and finally, the combined 5-PA versions of the 1D spectra. We anticipate these to become available in the latter part of the year 2017.

**Last Updated:**

This DOI is provided by the MAST archive at Space Telescope Science Institute

- [DOI FAQs](#)
- [DOI Main page](#)

- Closes the loop on Paper -> HLSP -> MAST -> Paper
- Better tracking statistics if DOIs are used in publications, esp. with support from, e.g. ADS.
- Naturally exposes HLSP to wide audience, expect increase in HLSP requests due to solicitation in paper submission process.





**MAST**  
**Users**  
**Group**  
**Meeting**

**January**  
**23–24,**  
**2018**

# Open Questions

1. How can MAST minimize work on contributors ***without*** sacrificing implementation of standards?
2. Why would an author NOT want to submit an HLSP to MAST? How can we mitigate those reasons, if MAST should at all?
3. How would you want to have HLSP show up in Portal searches when related to another MAST data set? What about those HLSP that complement a MAST mission (e.g., ground-based follow-up not directly tied to a single MAST mission file?)
4. How would you want to include HLSP catalogs in MAST services, either existing or new?
5. How would you want to find models and simulations provided as HLSP, both when tied to specific MAST observations and when not directly tied to real data products?



**MAST**

**Users  
Group  
Meeting**

**January  
23–24,  
2018**

# Science Impact of High-Level Science Products

(A Preliminary Report)

Dick Shaw

*Archive Sciences Branch  
STScI Operations and Engineering Division*





**MAST**

**Users**

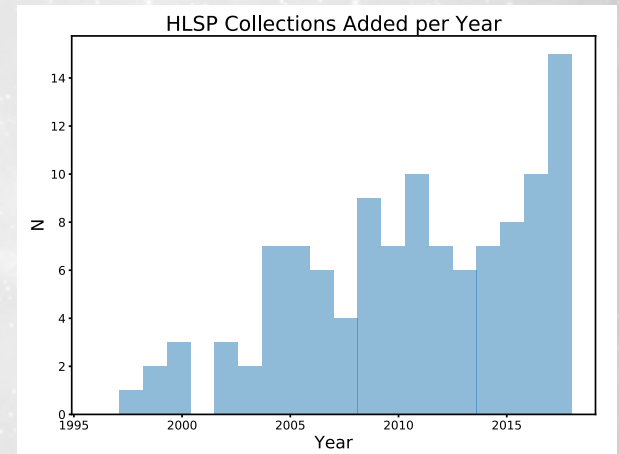
**Group  
Meeting**

**January  
23–24,  
2018**

# History of HLSPs

For 20 years MAST has solicited HLSPs from the astronomical community. They are contributed at the rate of several per year, in the form of collections of data and ancillary products that:

- Are related to a common science theme
- Have high potential to enable new research
- Are derived from, or are closely related to, MAST data holdings
- Have been used by the contributing team
- Are described in a refereed paper



Benefits to contributors & potential users include:

- Data hosted by MAST in a permanent archive; DOI assigned
- High quality data have been peer-reviewed & vetted by the community
- Products appear in Portal search results (or will soon)
- Collections featured in the MAST web presence



**MAST**  
**Users**  
**Group**  
**Meeting**

**January**  
**23–24,**  
**2018**

# HLSP Raison D'être

Why do we solicit and archive HLSP collections? The primary goal is to **enable new science** but there are a variety of supporting goals:

- Long-term preservation of high-quality, science-ready data
  - Personal/team websites usually disappear within a few years
- Overcome limitations of standard pipeline processing
- Preserve provenance
- Support reproducibility of science results
- Equalize access to high quality products
- Maximize science return for hosted missions

Are these goals are being achieved?

- How do we know that HLSP data are actually being used?
- What is the best/most appropriate measure of impact?
- Can we target new solicitations for HLSPs to maximize potential value?
- Can we better prepare HLSP collections to maximize re-use?



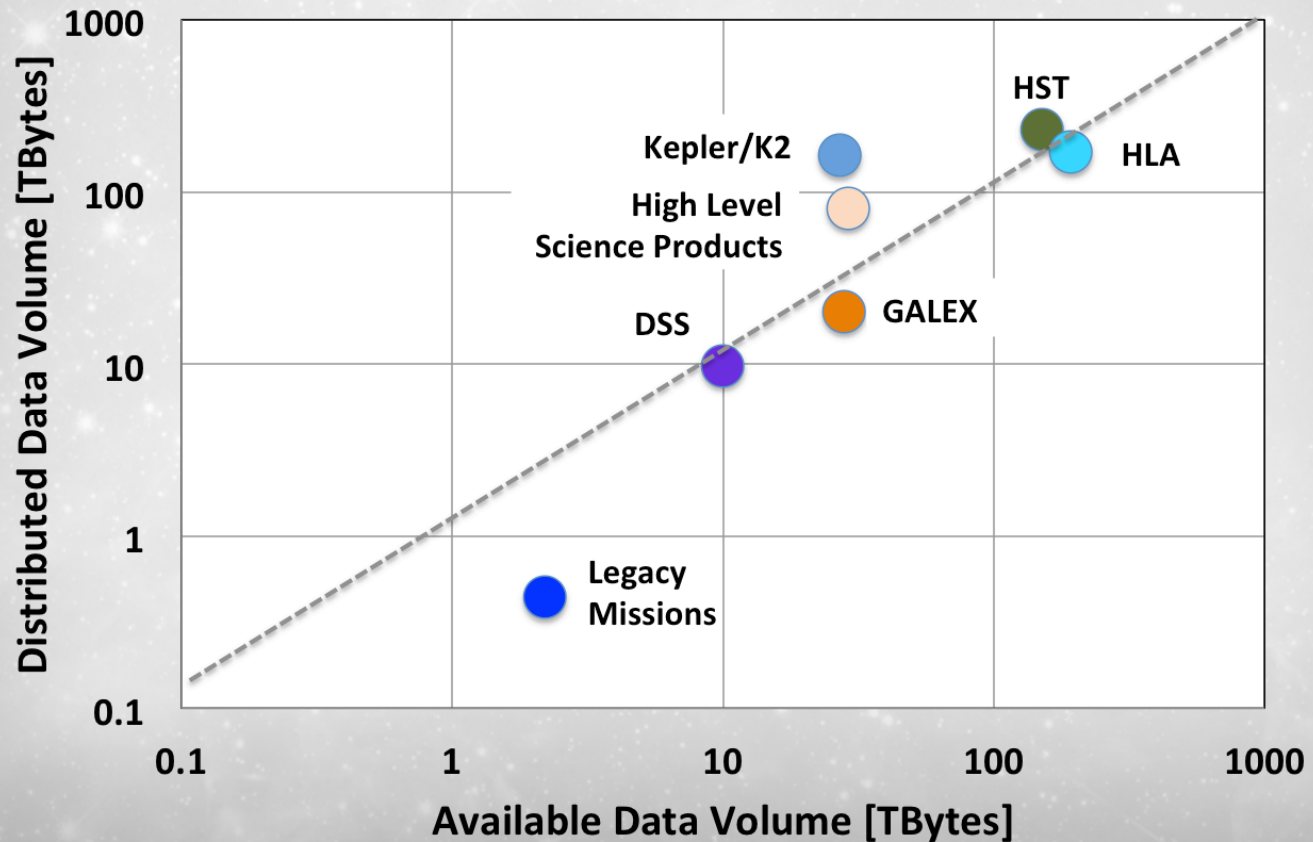


**MAST**  
Users  
Group  
Meeting

January  
23–24,  
2018

# MAST Data Retrievals

HSLP collections are typically more popular than standard data products from premiere NASA science missions.

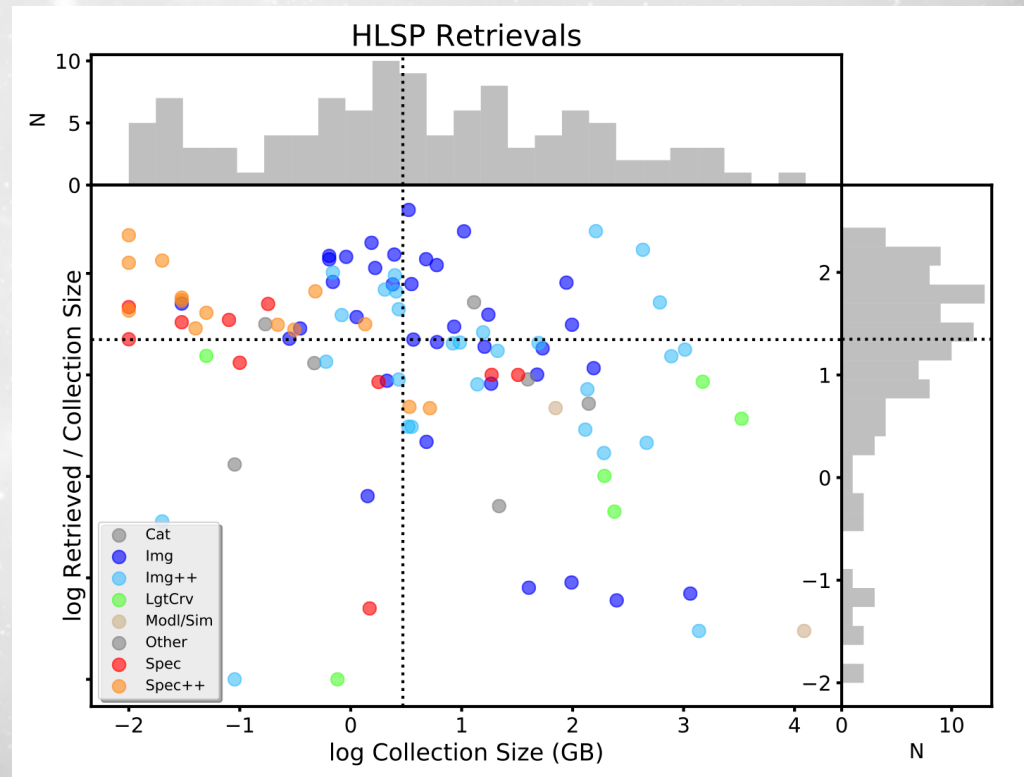


Retrieved data from 2016-Jan to 2018-Jan

# HLSP Data Retrievals

Considering just HSLP aggregate retrieval volume, in units of the collection size:

- Popularity is nearly independent of collection size
- Spectroscopic data are as popular as images and catalogs
- Lightcurves are now among the largest collections

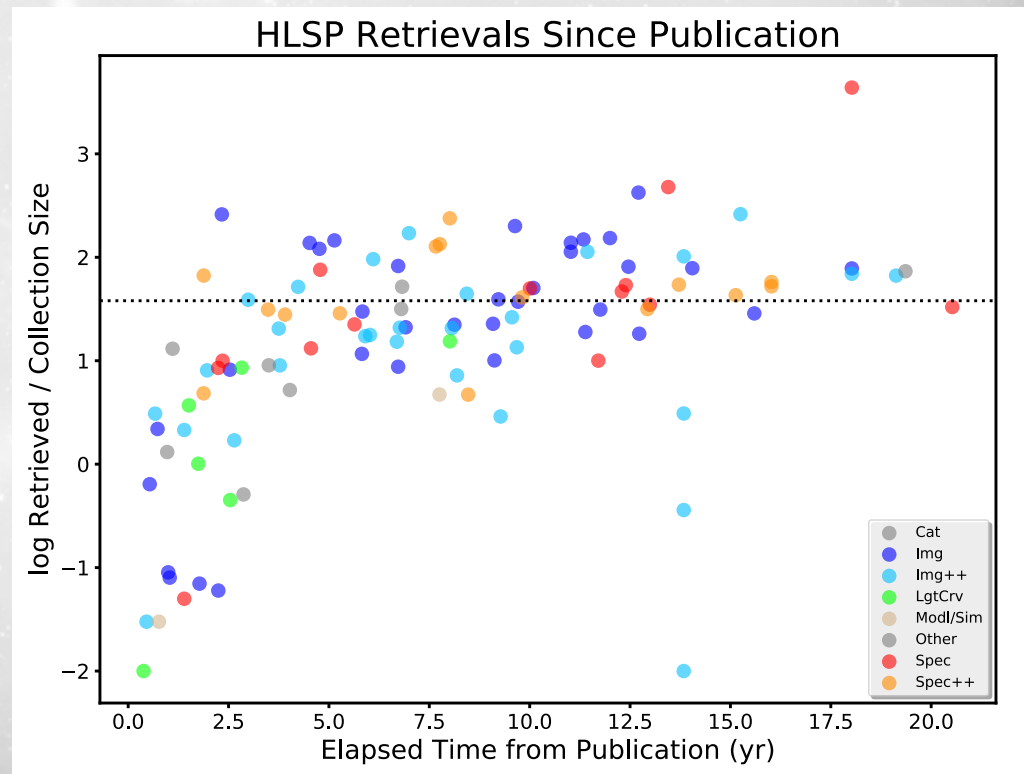




# HLSP Data Retrievals

Considering just HSLP aggregate retrieval volume, in units of the collection size:

- Popularity is nearly independent of collection size
- Spectroscopic data are as popular as images and catalogs
- Lightcurves are now among the largest collections
- It takes  $\sim 3$  yr for retrievals to accumulate

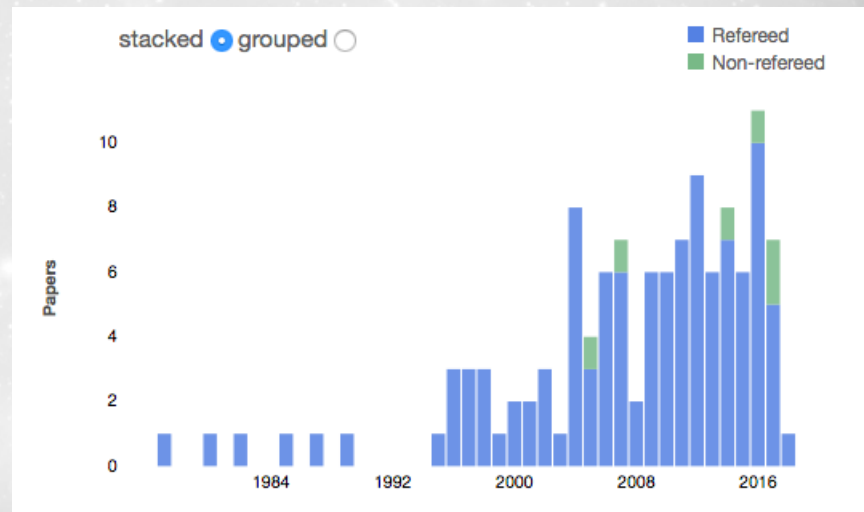




# HLSP Primary Papers

What is the science impact of HLSPs? A traditional approach is to measure citations, in our case to the HLSP *primary papers*\* in the refereed literature.

- Bibliometrics are generated from internal databases and **ADSBeta**
- 110/125 HLSP collections have a primary paper
- A few collections have more than one primary paper



Published primary papers per year. Graphics credit: ADSbeta.

See the ADS library **MAST HLSP** at [http://bit.ly/ads\\_hlsp](http://bit.ly/ads_hlsp)

\*Primary papers describe the HLSP data products and their creation.

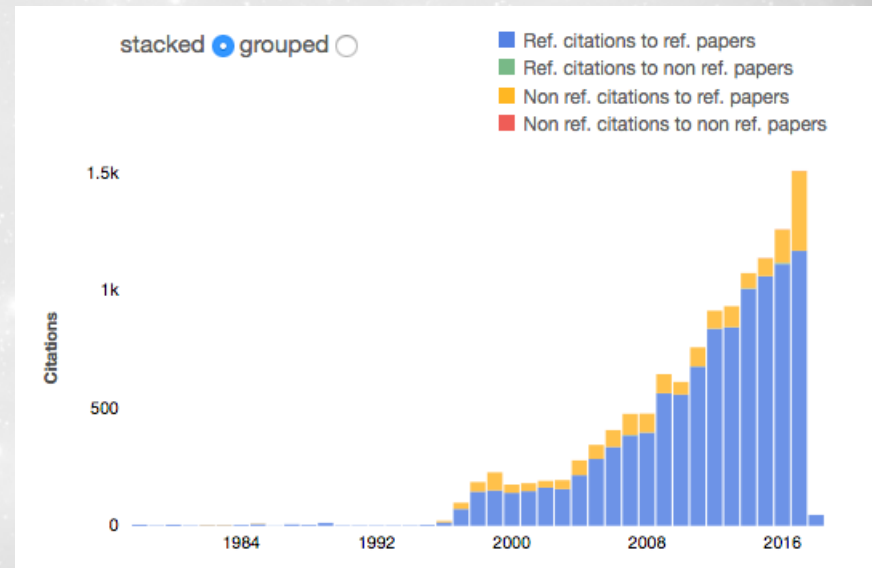




# HLSP Citations

Citation rates for HLSP primary papers are impressive. Considered as a collective body of work (122 papers):

- Nearly 12,900 citations from >9100 papers, and increasing steadily



Citations to HLSP primary papers per year. Graphics credit: ADSbeta.

See the ADS library **MAST HLSP** at [http://bit.ly/ads\\_hlsp](http://bit.ly/ads_hlsp)

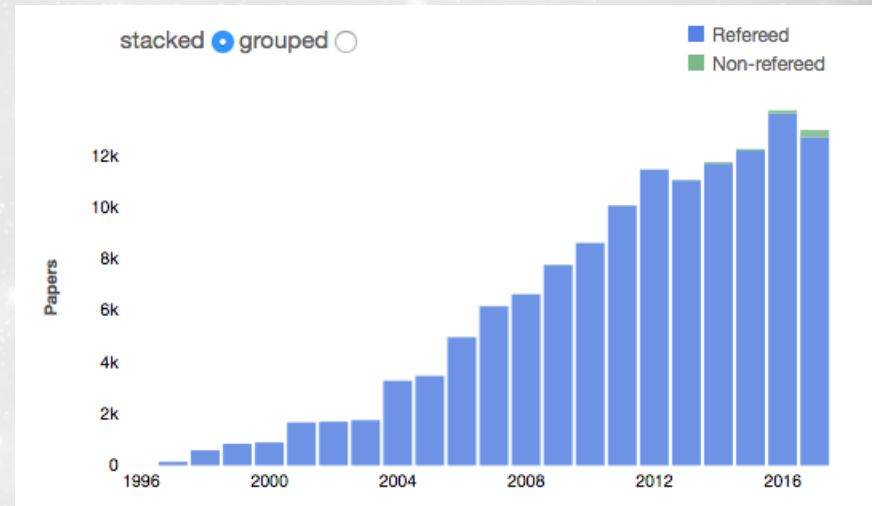


**MAST**  
**Users**  
**Group**  
**Meeting**  
**January**  
**23–24,**  
**2018**

# HLSP Citations

Citation rates for HLSP primary papers are impressive. Considered as a collective body of work (122 papers):

- Nearly 12,900 citations from >9100 papers, and increasing steadily
- Nearly 139,000 reads, and >79,000 downloads



Reads of HLSP primary papers per year. Graphics credit: ADSbeta.

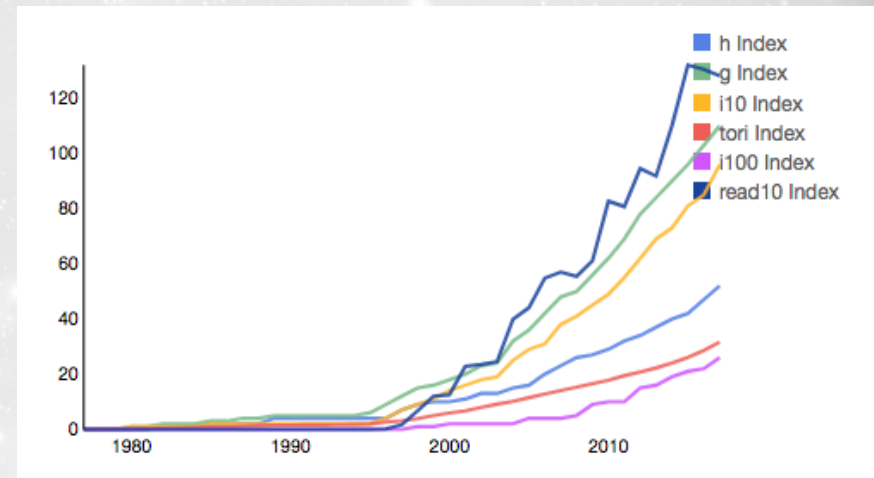
See the ADS library **MAST HLSP** at [http://bit.ly/ads\\_hlsp](http://bit.ly/ads_hlsp)



# HLSP Citations

Citation rates for HLSP primary papers are impressive. Considered as a collective body of work (122 papers):

- Nearly 12,900 citations from >9100 papers, and increasing steadily
- Nearly 139,000 reads, and >79,000 downloads
- H-index is 54, which compares favorably with other facility class missions



Various citation metrics per year. Graphics credit: ADSbeta.

See the ADS library **MAST HLSP** at [http://bit.ly/ads\\_hlsp](http://bit.ly/ads_hlsp)



**MAST**

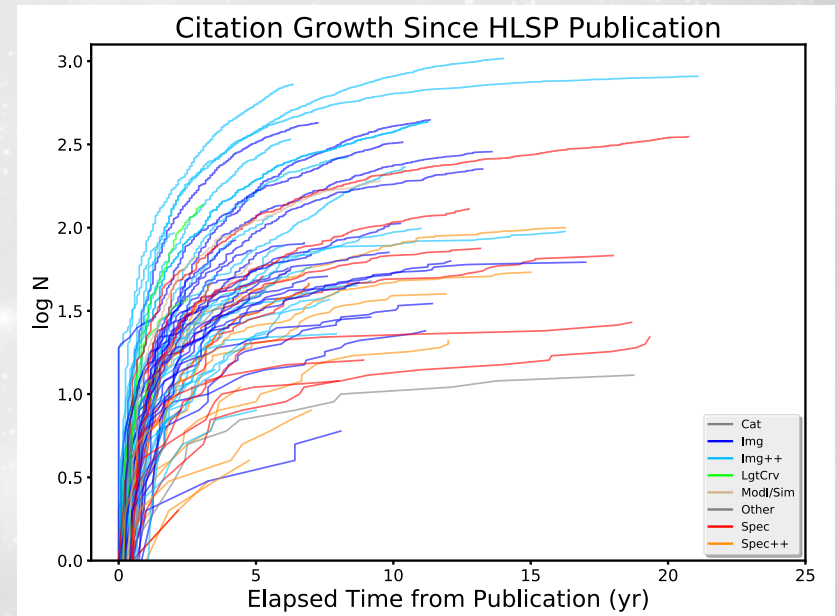
**Users  
Group  
Meeting**

**January  
23–24,  
2018**

# HLSP Citation Growth

Citations to primary papers accumulate quickly after publication, with a “time constant” of a few years.

- Citations continue to grow for decades after publication
- Most HLSP papers have >35 refereed citations (median for HST papers is 25)







MAST

Users

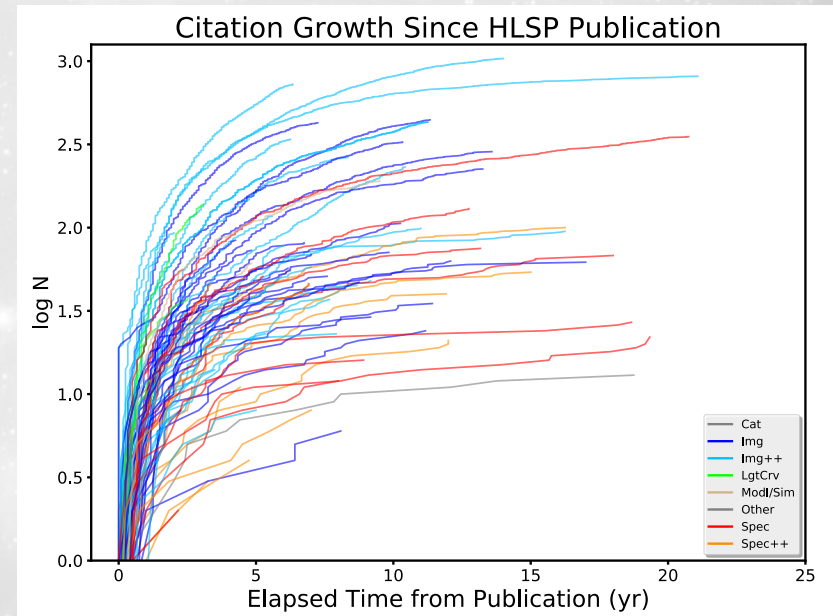
Group  
Meeting

January  
23–24,  
2018

# HLSP Citation Growth

Citations to primary papers accumulate quickly after publication, with a “time constant” of a few years.

- Citations continue to grow for decades after publication
- Most HLSP papers have >35 refereed citations (median for HST papers is 25)
- HLSPs containing image data products generally accumulate more citations
- The top 2 curves are the GOODS and HDF papers

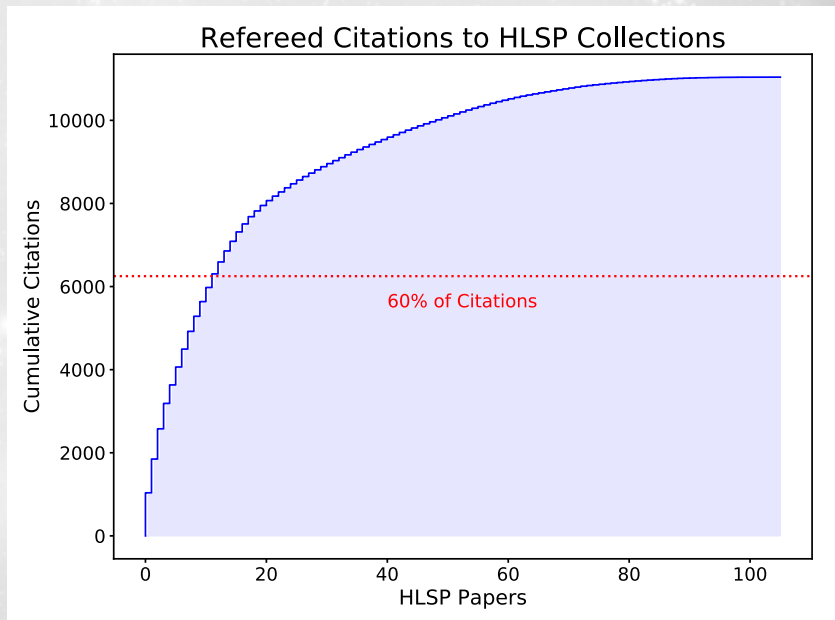




# HLSP Superstars

Citation rates vary considerably among the primary papers, for all the usual reasons.

- Primary papers in the top 10% account for 60% of all HLSP citations
- Of the 20 HST papers with >1000 citations (0.13%), 2 of them are HLSP primary papers
- Of the 209 HST papers with >300 citations (1.36%), 11 are HLSP primary papers



Cumulative citations





MAST

Users

Group  
Meeting

January  
23–24,  
2018

# Caveats

There are obvious shortcomings with measuring HLSP science impact solely on retrievals and citations.

- Hard to tell if HLSP data were actually used in citing papers
  - Are HLSPs a *driver* or a *by-product* of high-impact research?
  - This Data Science problem is being pursued by a Fellow from *InSight Data Science* (<http://insightdatascience.com/> in Seattle)
    - MUG opportunity: contribute toward a training set
  - Three HLSP primary paper examined in detail
  - One offering imaging data (37 citations):
    - ~75% of citing papers used the HLSP collection (images)
      - Of those, most were published soon after primary paper
    - One offering an HSC catalog (20 citations): 0 uses of data products
    - A third offering spectra & atlas (20 citations): 2 uses of data products
- Does not measure use of catalogs (e.g., w/CASjobs) very well
- Does not account for the *lack of correlation* between retrieval volume and citations
- Depends on authors doing a good job of citing data sources
  - DOIs for datasets will help, but have only recently been introduced



**MAST**

**Users**

**Group  
Meeting**

**January  
23–24,  
2018**

# Conclusions

It is clear that science programs which produce HLSPs often have a very high impact on science. The HLSP data are probably used a lot to enable new science.

- Bibliometrics are essential for quantifying the impact of HLSP collections
  - ADSbeta is already a very useful tool
  - Criteria for future HLSP acceptance must include a citable reference
- It is essential to know the extent to which HLSPs are actually used
  - DOIs for data collections will become an important tool
  - Full-text inference engine would be helpful (in progress)
  - Need a training set of 500-1000 papers
- MAST should
  - solicit HLSP collections early in the life of an observing program
  - publish the products as soon as possible once they are available
  - advertise their availability to the community
- It is important to raise community awareness of the value of contributing HLSP collections





**MAST**  
**Users**  
**Group**  
**Meeting**

**January**  
**23–24,**  
**2018**

# Questions for the MUG

- Can you suggest other good measures of HLSP science impact?
- What changes to HLSP solicitation, preparation, or curation would make them:
  - more valuable to the community?
  - have higher scientific impact?
- We are examining the question of whether citing papers made use of HLSP data.
  - Would HLSP curation still be worth continuing if the fraction of data re-use is: 25%, 10%, 0% ??
  - If some data product types routinely fail to attract interest, should MAST continue accepting them as HLSPs?
  - MUG opportunity: contribute to a substantial training set
- What can be done to attract new HLSPs with high potential?



**MAST**  
**Users**  
**Group**  
**Meeting**

**January**  
**23-24,**  
**2018**

# Backup Slides





**MAST**  
**Users**  
**Group**  
**Meeting**  
**January**  
**23-24,**  
**2018**

# HLSP Collections

HLSP Data Product Types by Volume

