Community-Contributed Datasets: Recent and Future Advances at the Mikulski Archive for Space Telescopes

Greg Snyder

Theresa Dower, Scott Fleming, Peter Forshay, Tim Kimball, David Rodriguez, Shui-Ay Tseng, Sarah Weissman

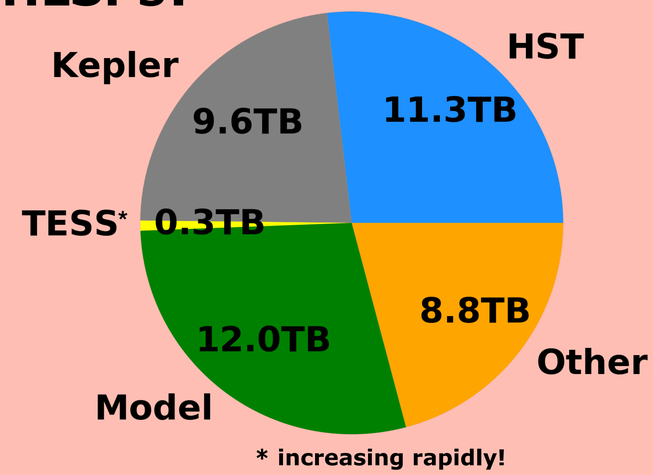
Space Telescope Science Institute

What are HLSPs?

- MAST High-Level Science Products (HLSPs) are science-ready data contributed back by community members
- Derived from or complementary to data from one or more MAST missions
- MAST offers permanent hosting, persistent identifiers (DOIs), and a variety of services

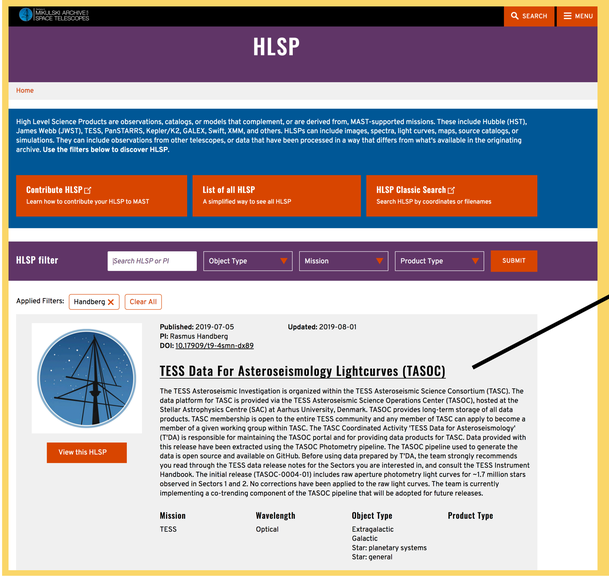
HLSP DOI example: See also: P2.17

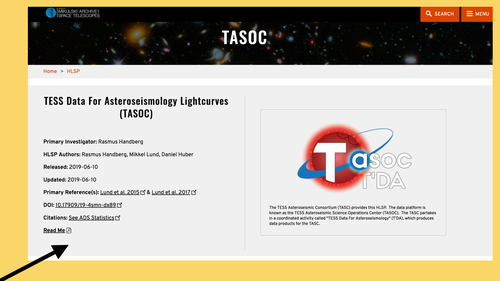
[1]The observed and rest-frame SEDs are available via DOI 10.17909/t9-3dbt-8734.



New search form and project pages

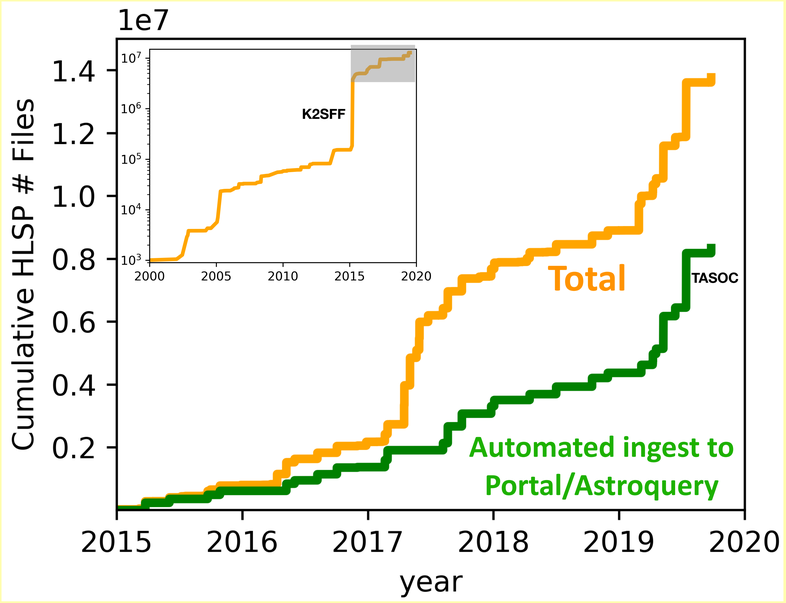
https://archive.stsci.edu/hlsp/





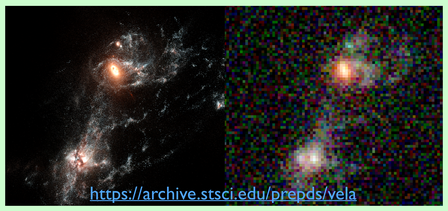
- Improved filtering based on HLSP categories
- Increased visibility of author names
- HLSP descriptions included on search page

Improve HLSP discovery in MAST services

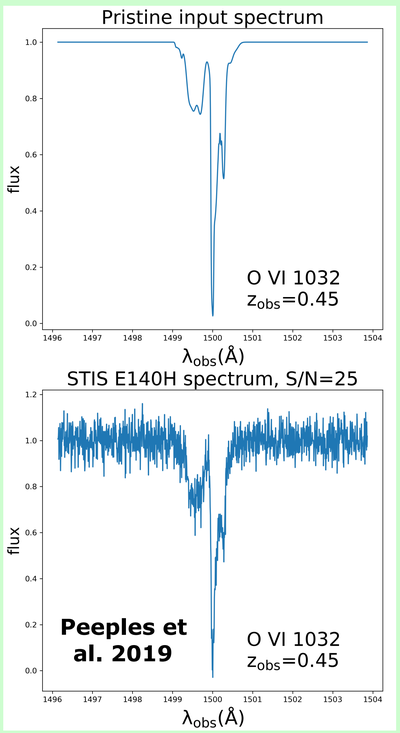


- Vast increase in # Files across the sky poses significant challenges to users
- New automated ingest to MAST services with code based on pyCAOM
- Makes files accessible through MAST Portal and Astroquery.MAST
  - So far: time-series data such as EVEREST, HALO, K2SFF, TASOC, K2SC, KEPSEISMIC, CDIPS
    - Already done > 3 million records, 60% of all HLSP files
- Focus in 2020: imaging surveys
- Other initiatives:
  1. VO registration for all HLSPs and TAP Schema for catalogs
  2. ASDI (see right)
  3. HLSP catalogs in MAST CASJobs

Enhance contact between models & data



- Models are 1/3 of HLSP holdings by data volume
  - Test theories
  - Calibrate complex observational signatures
- Heterogeneous, many-dimensional, and almost never map into database schema for observation products ("CAOM"): need custom services

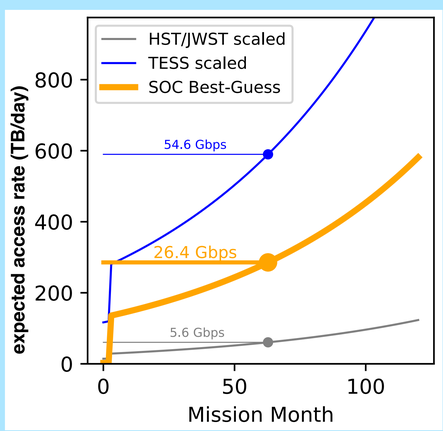


Archived Synthetic Data Initiative

- ASDI: An effort to host simulated datasets & unite them effectively with MAST holdings
- Prototype/Demo in progress for Jan. 2020
- Leverage STScI expertise with instruments and data simulation
- Expected Needs
  - Search and filter HLSP files not exactly fitting in data models
  - Enhance HLSPs of catalogs or catalogs + data products
  - Support new file types, e.g. HDF5
  - Standards and services for instrument simulation steps

Community products in the era of petabyte-scale surveys

- For WFIRST, MAST expects a 16 PB Science Archive derived from a few vast surveys
  - Exoplanet Microlensing: Expect demand similar to TESS
  - Large near-IR surveys: Similar to HST Treasury program HLSPs e.g. CANDELS
- Distributed Science Operations model: workflow similar to HLSPs
- Figure: using demand for HLSPs, TESS, and HST data, predict demand for WFIRST data access



Contribute!

http://archive.stsci.edu/hlsp/hlsp_guidelines/

Gregory F. Snyder

Astronomical Data Scientist

E: gsnyder@stsci.edu

References

- https://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/doc/pyCAOM/
- Novacescu et al. (2018), https://ui.adsabs.harvard.edu/abs/2018ApJS..236...20N
- Peeples et al. (2019), https://ui.adsabs.harvard.edu/abs/2019ApJ...873..129P
- Peeples et al. (Astro2020), https://ui.adsabs.harvard.edu/abs/2019arXiv190707184P
- MAST HLSP software: https://github.com/spacetelescope/MAST_HLSP
- MAST HLSP page: https://archive.stsci.edu/hlsp/