Preamble

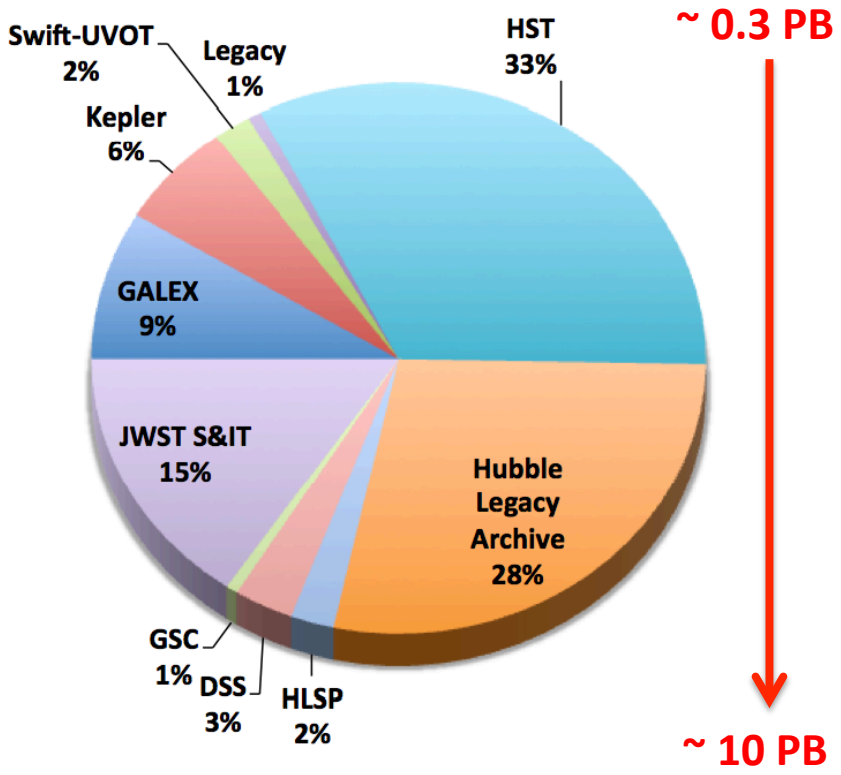
- Data-driven multi-wavelength science will play an increasing role in Astronomy over the next decade as large surveys like WFIRST/AFTA and LSST become realities.
- MAST, located at STScI, is a NASA funded archive to support and provide to the astronomical community a variety of astronomical data with the primary focus on related datasets in the optical, UV and NIR
- Within MAST a lot of attention devoted in the past to increase discovery, improve access, and create HLSPs for HST imaging, while HST spectroscopic HLSPs are presently very limited
- STScI recently embarked in effort to implement number of possible enhancements to HST archive to make spectroscopic data more useful to the scientific community

The Mikulski Archive for Space Telescopes @ STScI

FUV

MIR

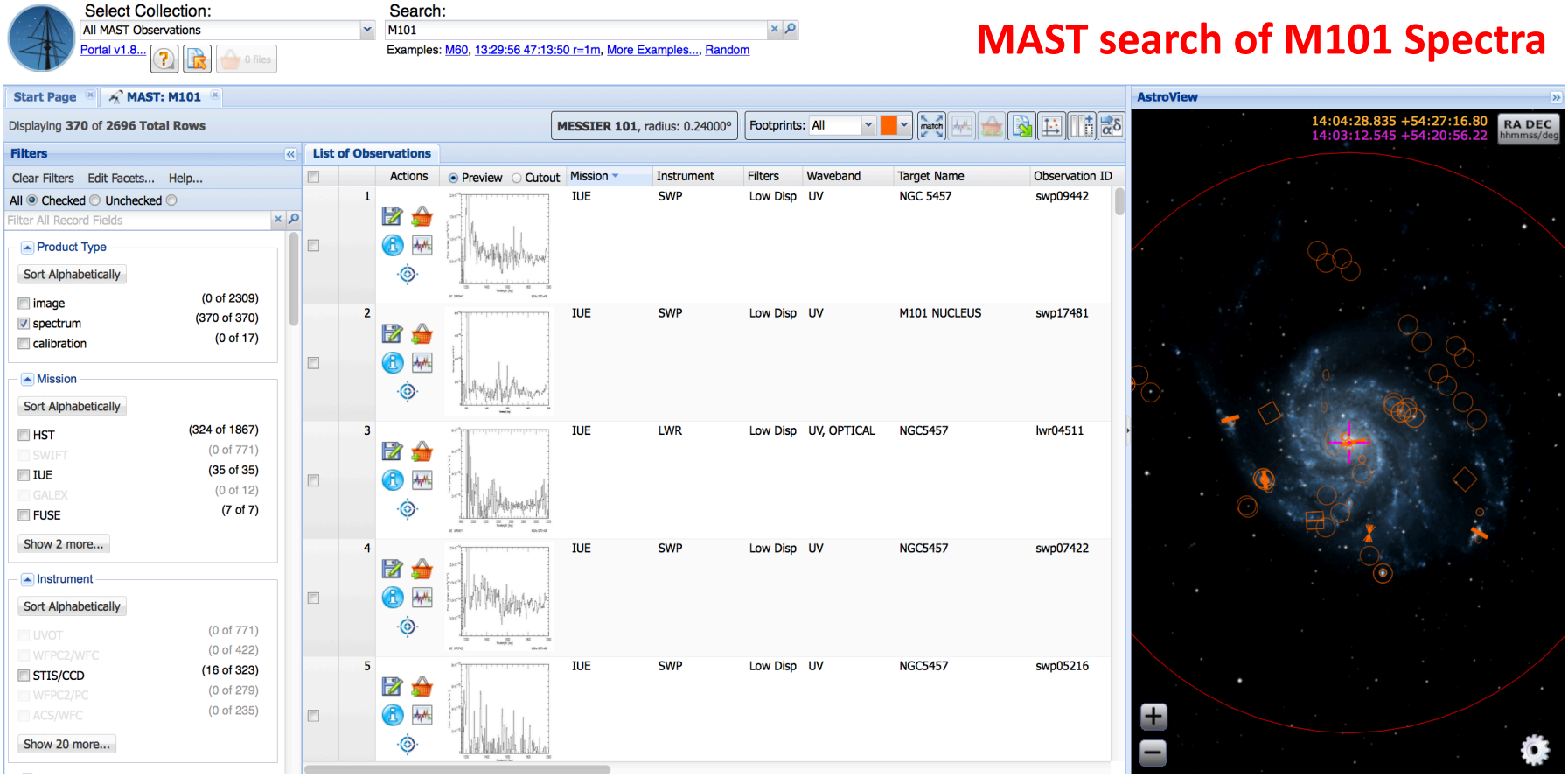
Mission/ Collection	Data Volume (GB)	Wavelength Range	Data Type
HST	107,230	0.1 – 2.2μ	I,Sp,sSp
HLA	90,000	0.1 – 2.2μ	I,Sp,sSp,Cat
Kepler/K2	20,066	4350 – 8450 Å	I,Cat,LC
XMM-OM	47	1500 – 6000 Å	I,Cat
HLSP	7505	70 Å – 2.2μ	I,Cat,Sp
SWIFT/UVOT	6641	1600 – 6000 Å	I
JWST I&T	51,000	0.6 – 28.5 μm	I,Sp
GALEX	28,590	1350 – 2800 Å	I,sSp,Cat
EPOCh	51	0.30 – 2.6 μm	I,Sp,LC
FUSE	1,200	905 – 1187 Å	Sp
IUE	600	1100 – 3200 Å	Sp
EUVE	96	70 – 760 Å	Sp
ASTRO	57	415 – 3300 Å	I,Sp
HPOL	0.2	0.32 – 1.05μ	Sp
ORFEUS	4.6	900 – 1400 Å	Sp
Copernicus	0.8	900 – 3150 Å	Sp
GSC2	2,500	4500 – 8500 Å	Cat
DSS	10,000	4500 – 8500 Å	I
VLA-FIRST	200	20 cm	I,Cat
TOTAL:	**318.0 TB**		



- MAST includes data from active (**RED**) and legacy (**BLACK)** missions covering the whole spectral range from FUV to MIR. Non-NASA funded projects are also included (**BLUE**).
- Data volume of current MAST holdings is ~ 300 TB. This will significantly grow into the PB scale over the next several years due to upcoming new missions that MAST will support.
- JWST data product mission 10-year baseline is ~ 1PB (not including working datasets or HLSP).
- MAST is playing or will be playing a significant role in the following ongoing/future missions: **PanSTARRs** (~ 2 PB), **TESS** (~ 20 TB), **WFIRST/AFTA** (3-9 PB), and **GAIA** (US-Mirror site; total of 300 TB by 2022).

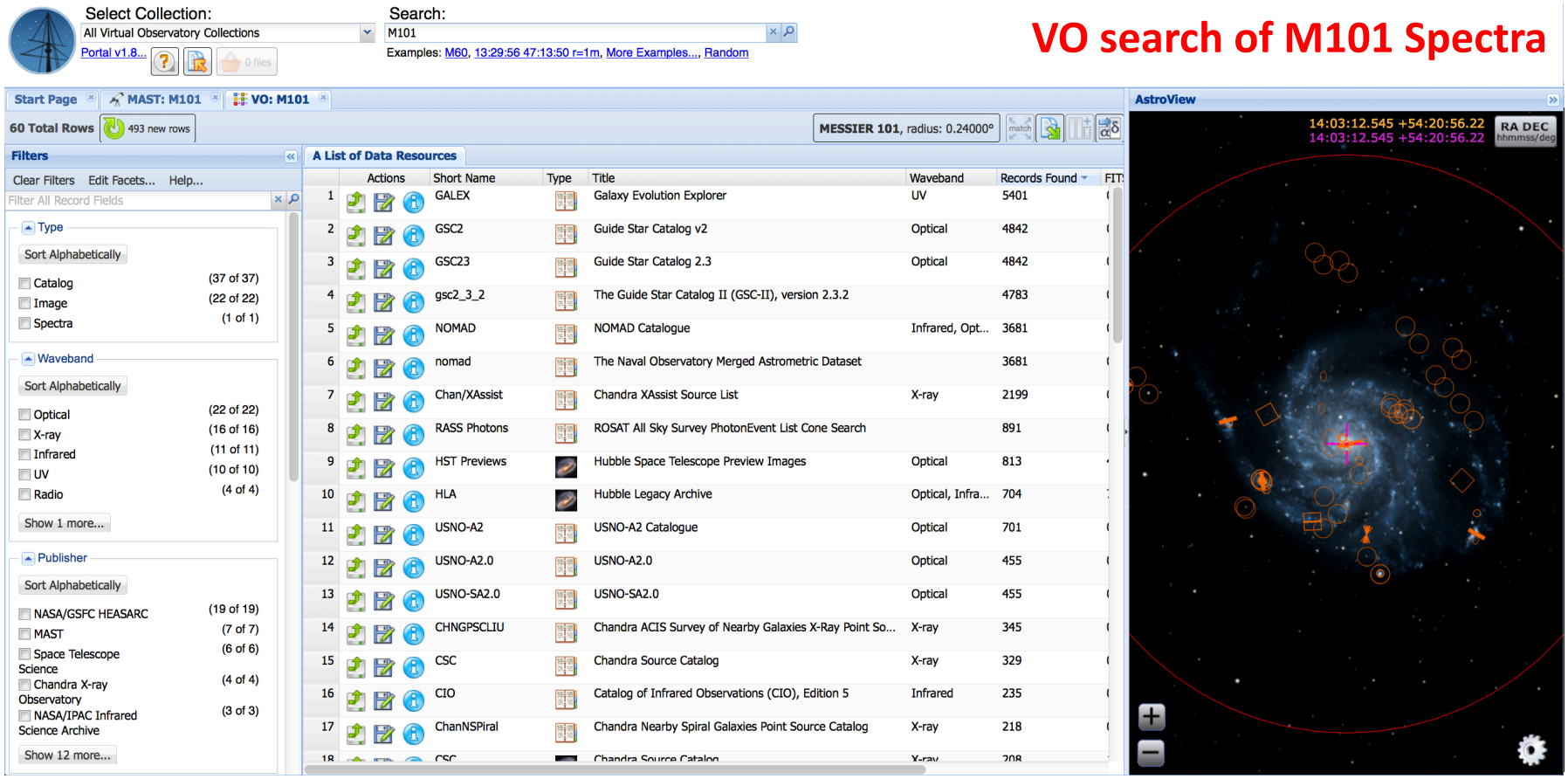
MAST Portal

1. Unify MAST missions with a common discovery interface.



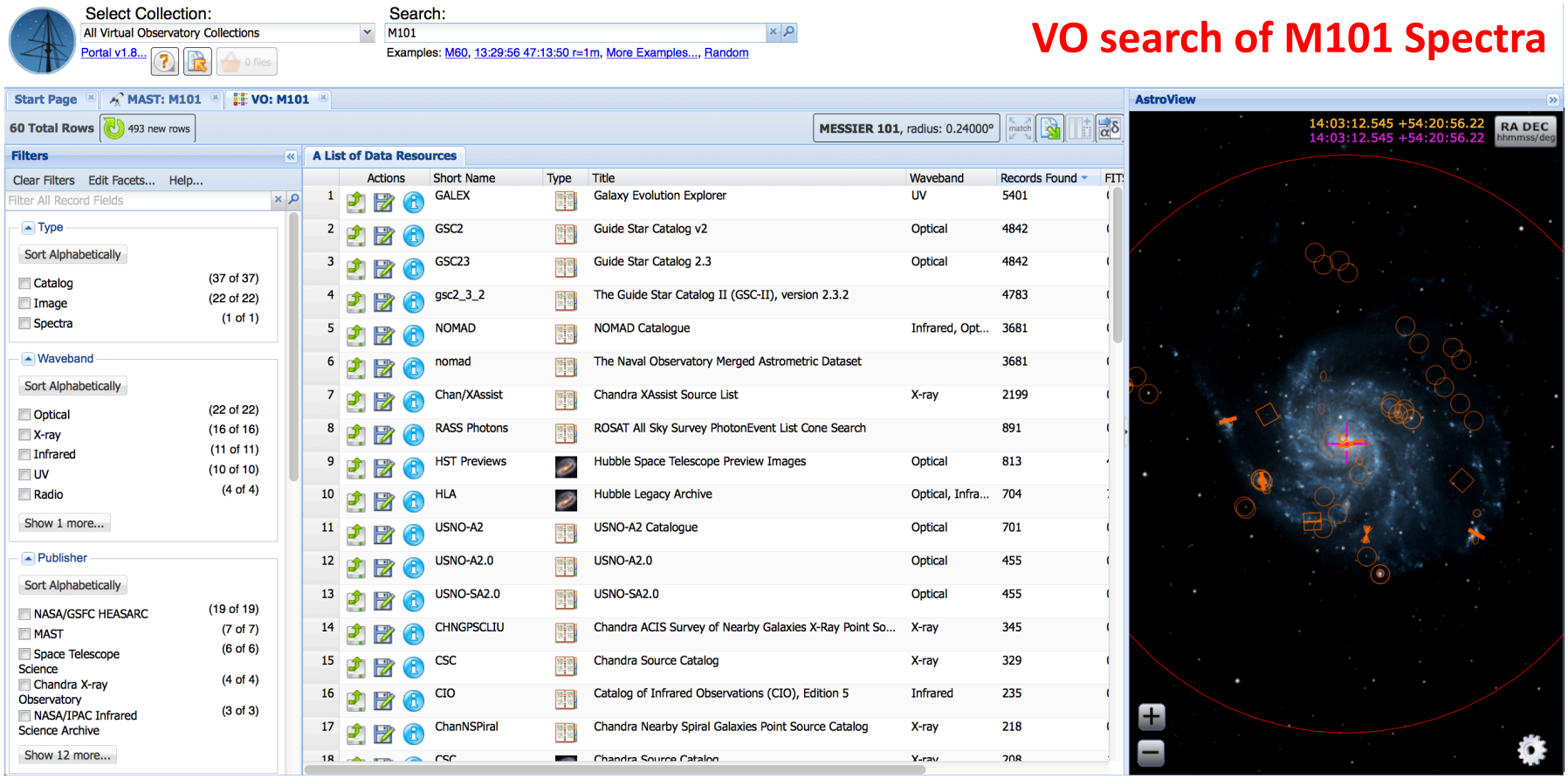
MAST Portal

1. Unify MAST missions with a common discovery interface.
2. Provide instant access to Virtual Observatory collections.

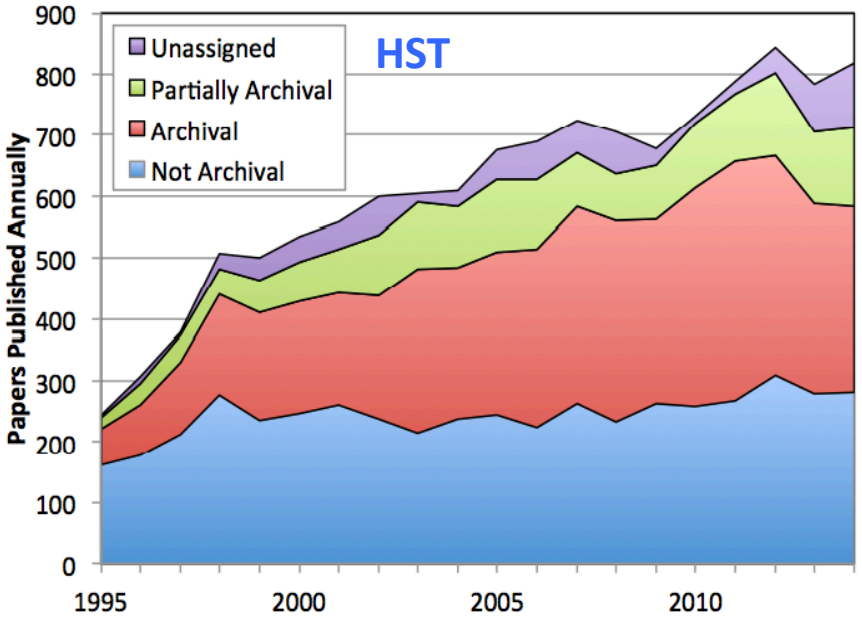


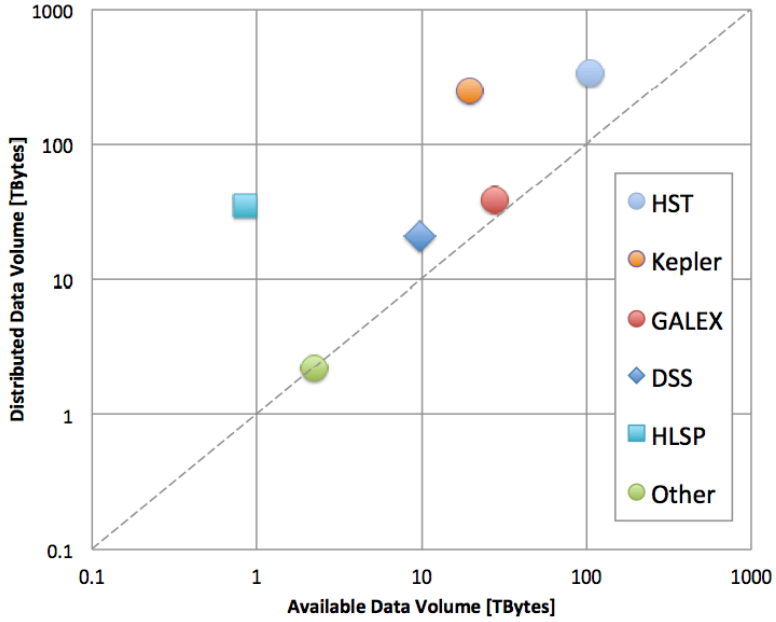
MAST Portal

1. Unify MAST missions with a common discovery interface.
2. Provide instant access to Virtual Observatory collections.
3. Build a framework for astronomy data interchange.



The Value of Archival Science and High-Level Science Products (HLSPs)





- 60% of MAST papers based, in whole or in part, on archival data! This is particularly true for HST.
- Missions with HLSP in MAST (e.g., HST and Kepler) show a distribution in volume ~ 2 mag higher than available data volume.
- Kepler HLSPs are mostly light curves.
- HST HLSPs are mostly imaging products produced by the community (e.g., the multi-cycle treasury programs) or the HLA project.

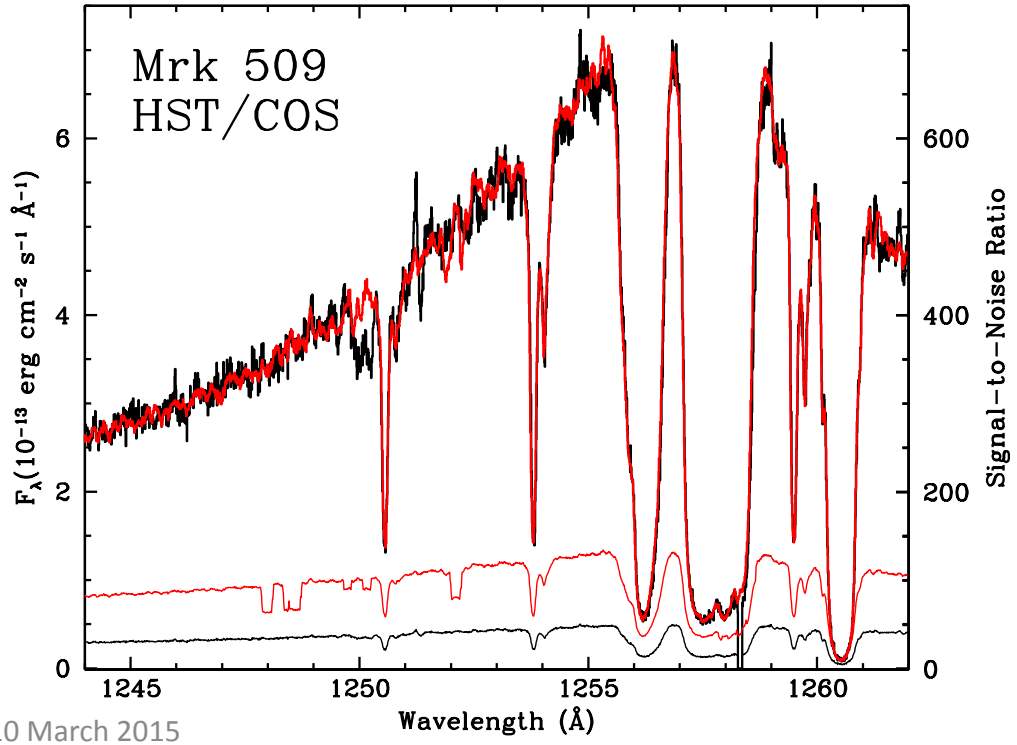
Need to focus on producing HLSP for spectroscopy !

Enhancing the Legacy of HST Spectroscopy

- Effort initiated a couple of years ago, following the Workshop entitled "Enhancing the Legacy of HST Spectroscopy" held at STScI on Nov 15-16, 2012.
- Primary goals of this effort include:
  - ✓ Optimization of HST impact on current and future spectroscopic research (both new and archival) by identifying the most productive and archival improvements from the suggestions generated by the Workshop
  - ✓ Ensuring that the HST final archive will have the quality, content, and tools to support needs of the science community until next UV/optical spectroscopic mission is launched, a decade or more in the future.
- 5 major areas were identified:
  - ✓ Spectral combination
  - ✓ Consolidation of spectral visualization tools into the portal
  - ✓ Target classification
  - ✓ Extraction of information from spectra
  - ✓ Automatic extraction of multi-object spectra

Spectral Combination and definition of new HLSPs

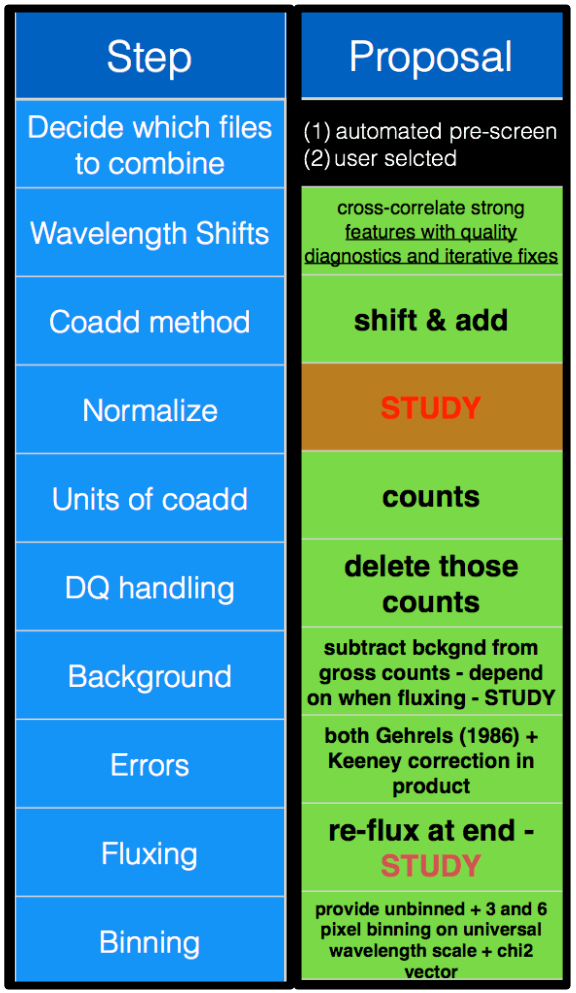
- Ultimate purpose: combine 1D spectroscopic data and create new HLSPs.
- Effort can be thought as equivalent of *ASTRODRIZZLE* for spectroscopic data.
- An example of advantage of combining spectra taken, e.g, with different COS settings is illustrated below.



Combined HST/COS spectrum of the Seyfert 1 galaxy Mrk 509 from Kriss et al. (2011, A&A, 534, 41). The black curve shows the single spectrum of one wavelength setting produced by the COS pipeline. The red curve shows the combined spectrum using all four central wavelength settings (this is not currently done by CALCOS). The thinner curves show the respective signal-to-noise ratios of the single spectrum and the combined spectrum.

Spectral Combination and definition of new HLSPs (cont.)

- HST Spectroscopic Data Products Working Group created to build a prototype pipeline
- Writing of the pipeline envisioned in steps:
  - ✓ will start as off-line tool manually triggered to create new HLSPs for HST data into the HLA
  - ✓ could evolve into part of the automatic DMS processing
  - ✓ could become a stand-alone tool offered to community, similarly to an *ASTRODRIZZLE* package for spectroscopy
- Effort done incrementally by increasing difficulty and complexity of 1D spectral data
  - ✓ Starting with FUV COS and FUV+NUV STIS Echelle data
- Several pipelines developed by individuals/science teams used to facilitate design and requirements.
  - ✓ Existing pipelines tailored to specific science goals.
  - ✓ A more general design required.
- Steps for prototype pipeline identified.
- In the process of finalizing the prototype.

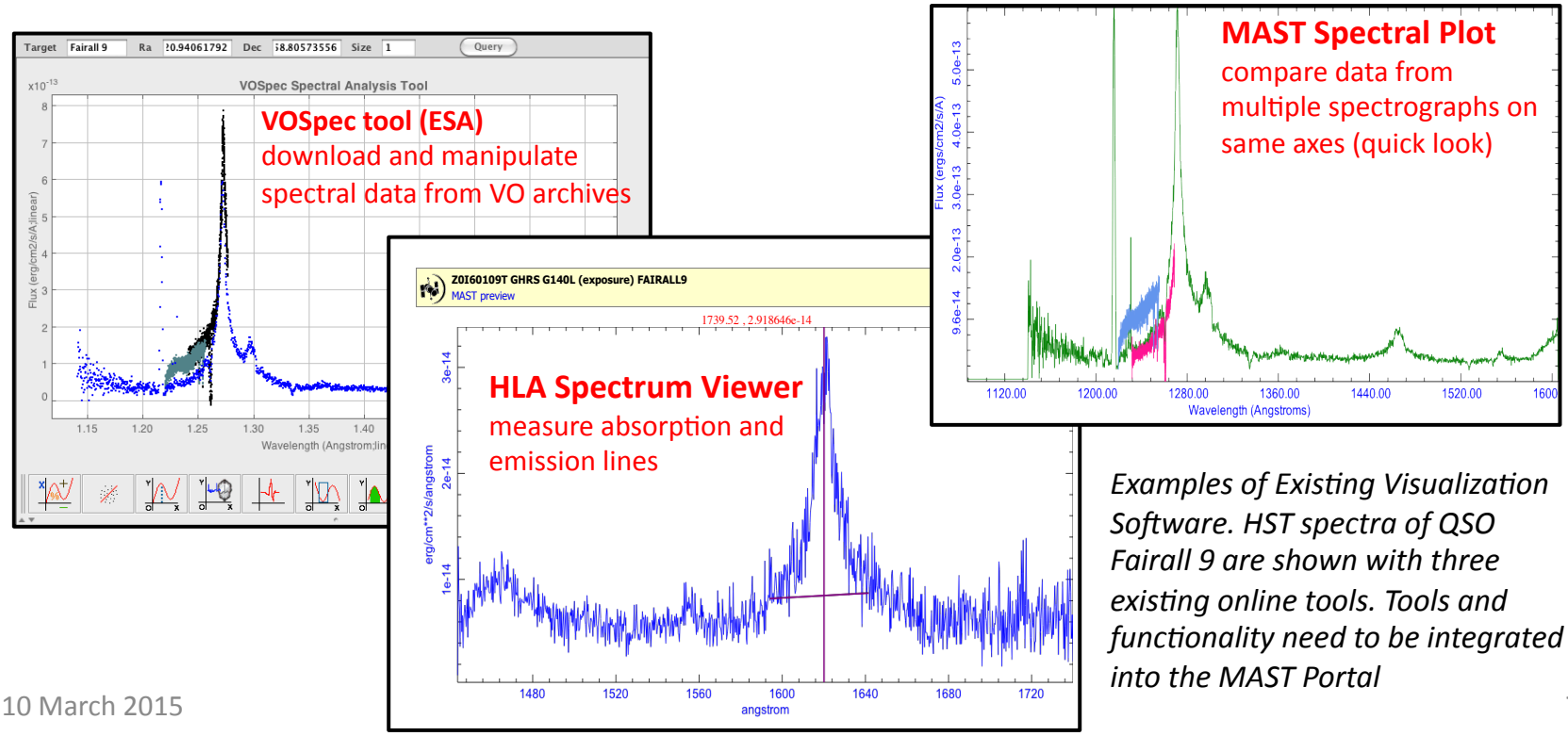


Step	Proposal
Decide which files to combine	(1) automated pre-screen (2) user selcted
Wavelength Shifts	cross-correlate strong features with quality diagnostics and iterative fixes
Coadd method	**shift & add**
Normalize	**STUDY**
Units of coadd	**counts**
DQ handling	**delete those counts**
Background	**subtract bckgnd from gross counts - depend on when fluxing - STUDY**
Errors	**both Gehrels (1986) + Keeney correction in product**
Fluxing	**re-flux at end - STUDY**
Binning	**provide unbinned + 3 and 6 pixel binning on universal wavelength scale + chi2 vector**

Courtesy J. Tumlinson (STScI)

Consolidation of Spectral Visualization Tools into the MAST Portal

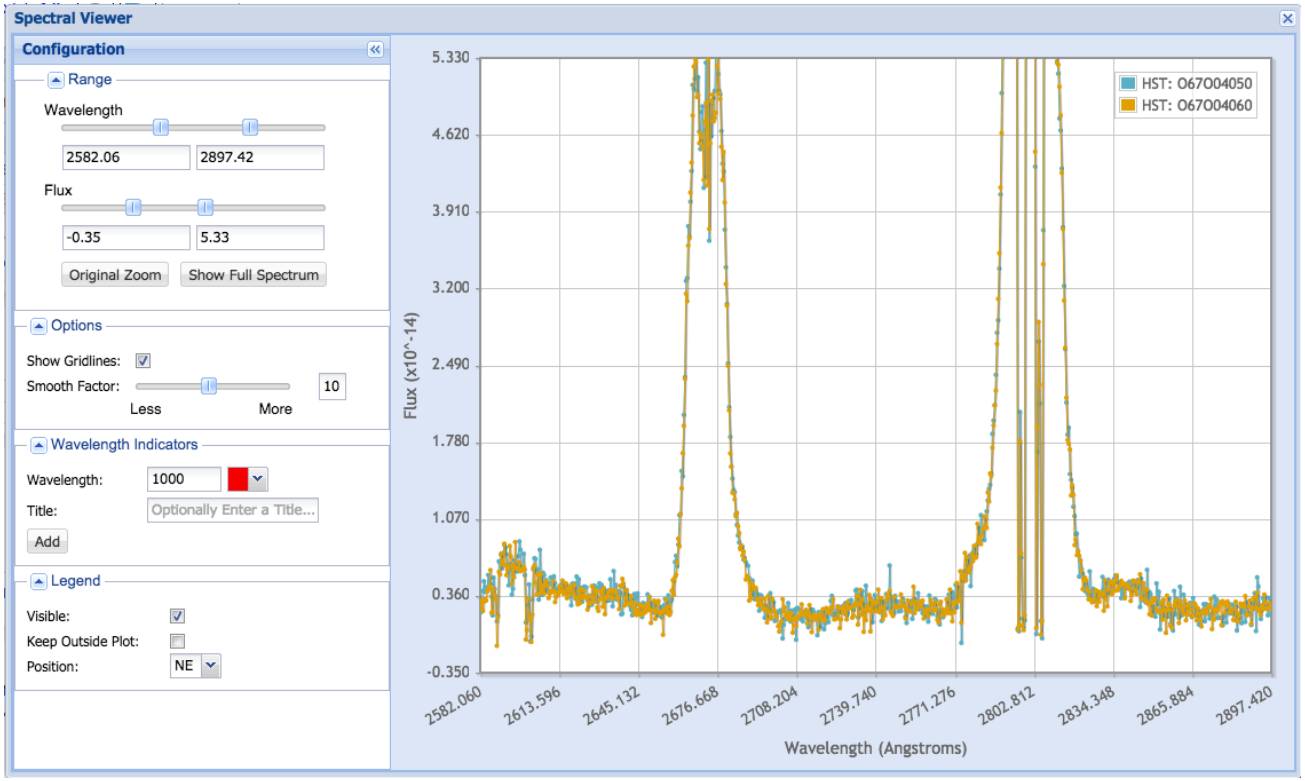
- Tools with different functionality already exist but spread throughout MAST, HLA, and VO (see figures below):
  - ✓ OK for quick look and some basic measurements
  - ✓ Need improvements to become more user friendly and for more in depth analysis
  - ✓ Should be consolidated into the MAST Portal



Examples of Existing Visualization Software. HST spectra of QSO Fairall 9 are shown with three existing online tools. Tools and functionality need to be integrated into the MAST Portal

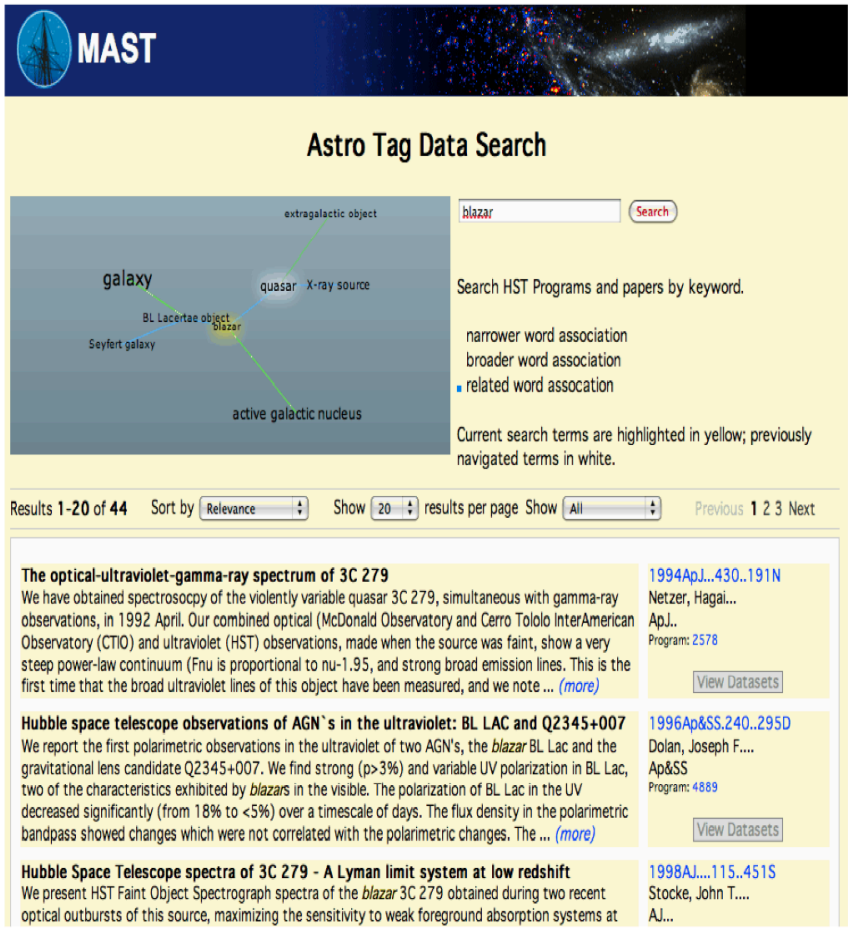
Consolidation of Spectral Visualization Tools into the MAST Portal (cont.)

- New interactive web-based spectral viewer with some of the desired features released into the MAST Portal in Feb 2015
- Will allow for quick look and basic measurements (e.g., line fitting) of spectral data available in the archive



Target Classification

- Main focus of effort is creation of dynamic search tool integrated into the MAST Portal that will select targets based on some kind of classification
  - ✓ applicable to all HST data (imaging and spectroscopy)
- Current problem is that HST sources are not properly classified with conventional lexicon during Phase I and/or Phase II.
- ***AstroTag*** project in progress
  - ✓ Semantic tagging of data
  - ✓ consists in associating datasets to keywords by using the Unified Astronomy Thesaurus (UAT, astrothesaurus.org; maintained by AAS)
  - ✓ Will provide alternative way to access data based on science topics

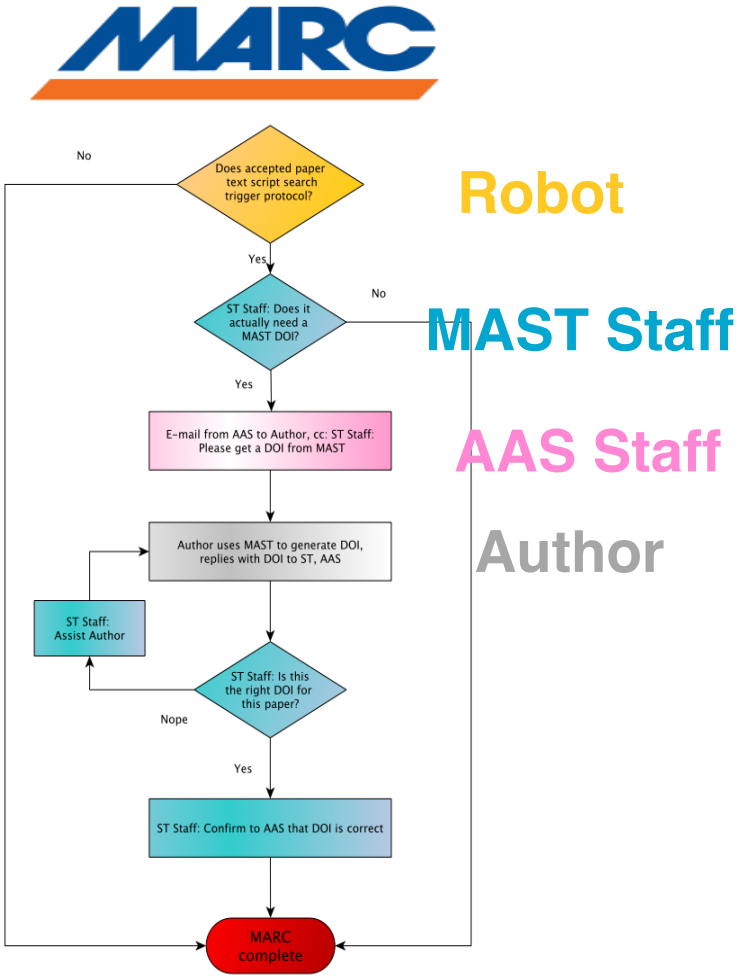


Prototype Interface for exploring data based on semantic tagging. Courtesy S. Weissman and MAST developers (STScI)

Target Classification (cont.)

Classification of sources approached incrementally

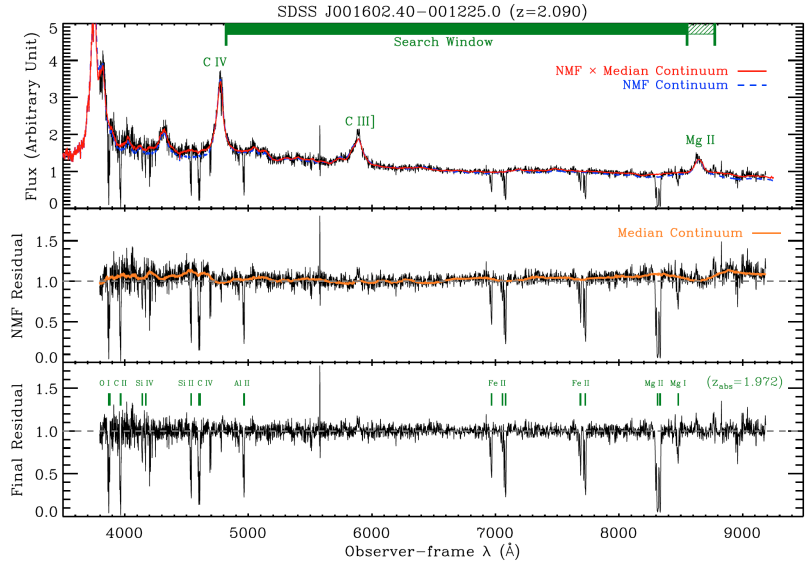
- HST program IDs (PI keywords & proposal abstracts) already associated to literature
  - ✓ Keywords could be easily extracted for each source in the program, even if process not quite rigorous and subject to false positives
- HST datasets (in addition to program IDs, PI keywords & proposal abstracts) are currently being associated to literature
  - ✓ At present done manually (resource intensive)
  - ✓ ***MARC*** initiative in progress: DOI (a broadly used, permanent, citable URL) will be used to connect MAST datasets to AAS journals
- Future developments could include catalog-based classification, crowd-sourced classification, better definition of target keywords in Phase II



Courtesy J. Peek (STScI)

Extraction of Information from Spectra

- Purpose is to develop tools for feature identification and measurements on spectra
  - ✓ Tools would be integrated into the MAST Portal
- Effort can be thought as equivalent of *Source Extractor* for spectra
- Similar to what already done, e.g., for absorption lines in quasars of the SDSS:

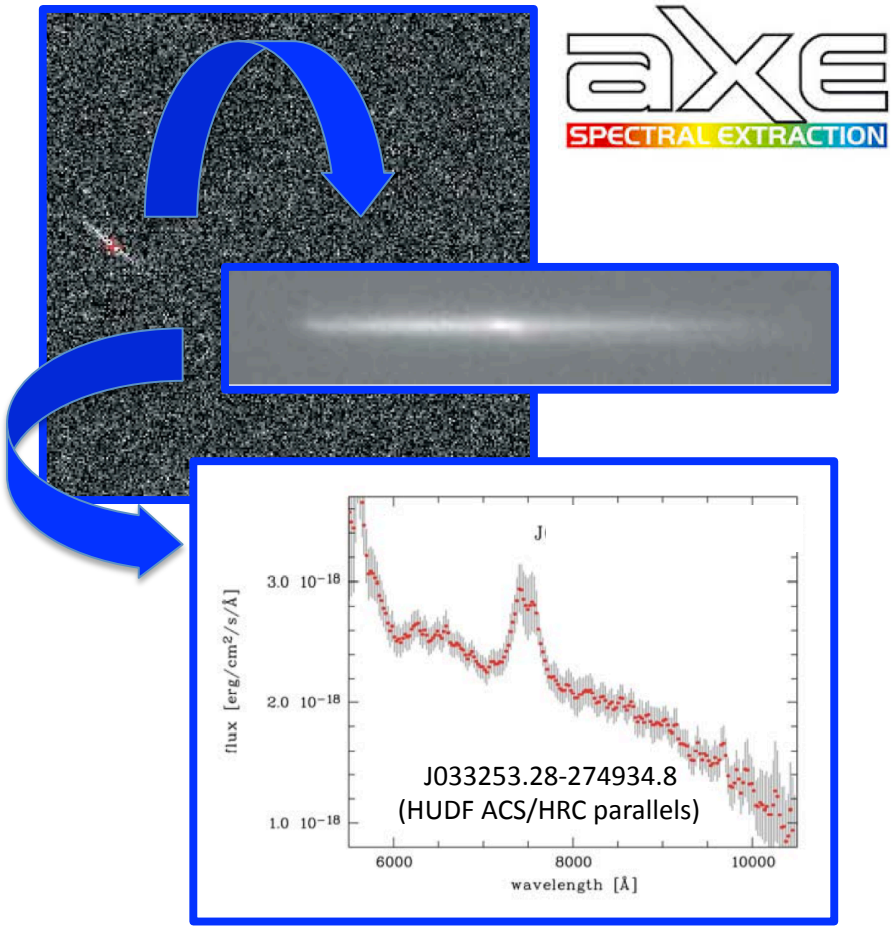


Automated fitting of quasar spectrum with automated detection of absorption lines (Fig. 1 from Zhu & Menard 2013, ApJ, 770, 130). Top panel shows best fit nonnegative matrix factorization (NMF) quasar model from SDSS DR7 plotted on observed spectrum. Dividing this out leaves a normalized spectrum with broad residual structures (2nd panel). Median filtering this spectrum produces bottom panel, which is then searched automatically for absorption lines.

- It requires multiple options to handle different types of data (emission vs absorption, bands, forests; high S/N vs low S/N; high *z* v. low *z*, continuum structure)
  - ✓ relies on classification of the spectroscopic source
- Effort not yet started

Automatic Extraction of Multi-Object Spectra

- Currently only limited MOS datasets in MAST/HLA (from ad-hoc extraction of HST/ACS grism spectra by ST-ECF)
- HST+JWST+WFIRST working Group at STScI currently investigating new algorithms to perform optimal extraction of MOS spectra
  - ✓ Document expected by Fall 2016
- Effort may lead to the re-write of the IRAF aXe software package in Python or the creation of a new open-source software package more flexible and mission independent



Archive Strategic Roadmap @ STScI

- Data Management Plan at STScI currently undergoing an effort to update the strategic roadmap for the Archive over the next 3-5 years
- Items of interest for the Astronomical community include:
  - **Multi-Mission Operations** (*scalable* architecture with distributed workflow processing and networked storage solutions, allows for affordability of small missions)
  - **Partnerships with other Archives** (common data interchange *models* e.g. CAOM, VO standards via the NASA-VO collaboration)
  - **Collaborative Resources** (VMs for science users, science cloud, *Integrated Science Services* for the Observer)
  - **Data Discovery and Data Mining** (3D data visualization, non-positional search, *indexing* and *algorithms* to discover features in data)
  - **Data analysis** (*open-source tools* to perform scientific analysis of the data on-line, including use of catalogs, e.g., HSC, to create and interpret diagrams or select sources based on their photometric properties, photometric redshifts, etc.)

MAST Contacts @ STScI

Alessandra Aloisi (DMS Program Manager)
Gretchen Greene (DMS Chief Engineer)
Karen Levay (Archive Sciences Branch Head)
John MacKenty (Grisms WG chair)
Joshua Peek (MARC/DOI initiative lead)
Marc Postman (Community Mission Office Head)
Jason Tumlinson (HST Spectroscopic Data Products WG chair)
Sarah Weissman (AstroTag project lead)
Rick White (MAST PI)