

The STScI Archive and Data Center

Overview of current and planned activities

Rick White

STUC

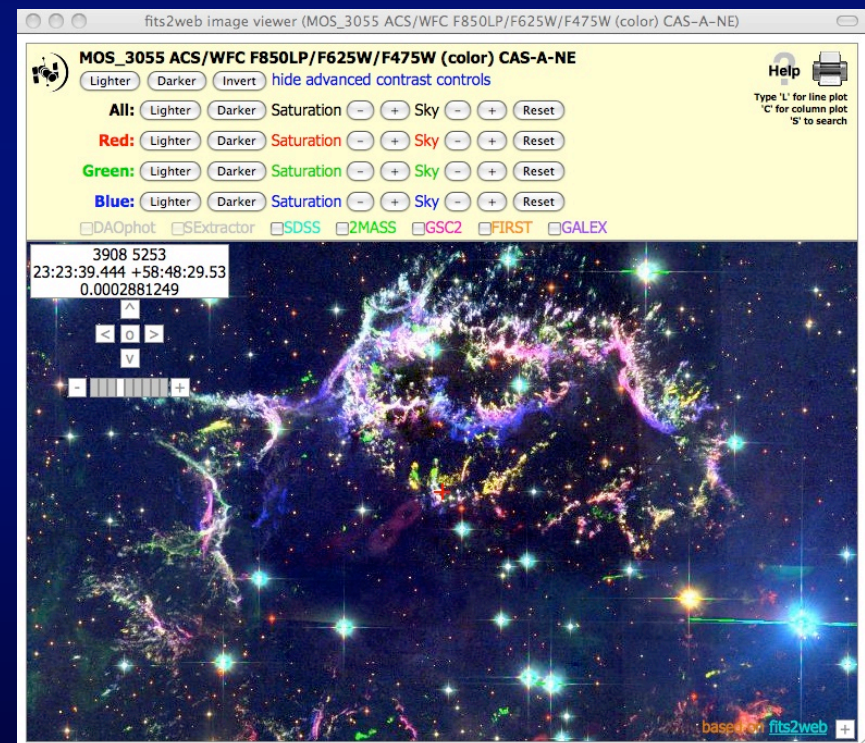
2010 April 13

The STScI Archive and Data Center



- Overview of the STScI archive and missions
- Science impact of the archives
- History of the HST archive
- Selected recent activities
- Hubble Legacy Archive: DR4
- Near-term plans

Cas A
HLA ACS mosaic



STScI Archive and Data Center



- STScI hosts archives and data processing for multiple missions
 - The big active missions: HST, Kepler, GALEX
- HST archive has been in operation since Hubble launch in 1990
 - All HST data are retrieved through the archive
 - Hubble Legacy Archive with enhanced data products open since 2008
- MAST: Multi-mission Archive at Space Telescope
 - Established in 1997 as NASA's Optical/UV archive
 - Supports both active and legacy missions (IUE, FUSE, EUVE, ...)
 - The other NASA archive centers:
 - HEASARC (GSFC): X-ray, gamma ray
 - IRSA/IPAC/NED/... (Caltech/JPL): Infrared
 - ADS (CfA): Astronomical literature

What is the STScI archive?



- Data

~125 TB of images, spectra, catalogs, time series

- Metadata

~10⁶ HST observations (plus other missions)

Documentation, publication links, ...



Hubble Legacy Archive

m51
Examples: M101, 14 03 12.6 +54 20 56.7 +/-0.2d, more...
Requires Firefox, Safari, IE7, or compatible browser

Inventory Images Footprints Cart, 0 kB Grism Spectra (ST-ECF) Help

m51 RA = 202.482194 Dec = 47.231509 r = 0.075000 [13:29:55.727 +47:13:53.43]

Results 1-20 of 224 Show 20 results per page Previous 1 2 3 4 5 6 7 8 9 ... 12 Next

Click column heading to sort list - Click rows to select Add selection to cart
Show selected rows: First Mixed Only Not Reset selection
Text boxes under columns select matching rows Apply Filter Clear Filter

Display	Retrieve	RA	DEC	Level	Target	Detector	Aperture	Spectral_Elt
Display	<input checked="" type="checkbox"/> FITS	13:30:07.45	47:16:11.3	5	M51-POS6	ACS/WFC	WFCENTER	F435W
Display	<input checked="" type="checkbox"/> FITS	13:30:07.45	47:16:11.3	5	M51-POS6	ACS/WFC	WFCENTER	F555W
Display	<input checked="" type="checkbox"/> FITS	13:30:07.45	47:16:11.3	5	M51-POS6	ACS/WFC	WFCENTER	F658N
Display	<input checked="" type="checkbox"/> FITS	13:30:07.45	47:16:11.3	5	M51-POS6	ACS/WFC	WFCENTER	F814W

Instrument #Footprints

<input checked="" type="checkbox"/> ACS	184
<input checked="" type="checkbox"/> ACSGrism	0
<input checked="" type="checkbox"/> WFPC2-PC	56
<input checked="" type="checkbox"/> WFPC2	83
<input checked="" type="checkbox"/> STIS	30
<input checked="" type="checkbox"/> NICMOS	55
<input checked="" type="checkbox"/> NICGrism	0
<input checked="" type="checkbox"/> FOS	15
<input checked="" type="checkbox"/> GHRS	0

DSS Image ☐ On ☐ Off

Data Product

☐ Exposure (Level 1)
☐ Combined (Level 2)
☐ Best Available
☐ Contributed HLSP
☐ Mosaic (Level 3)

Submit

To Zoom, go to Advanced Search and enter a smaller value for Drizzle fractional width 0.11

```
<TABLE>
<DESCRIPTION>STScI Hubble Legacy Archive SIAP</DESCRIPTION>
<INFO name="QUERY_STATUS" value="OK"></INFO>
<RESOURCE type="results">
  <PARAM datatype="char" name="INPUT:POS" value="210.802458,54">
  <PARAM datatype="double" name="INPUT:SIZE" value="0.240000">
  <PARAM datatype="char" name="INPUT:FORMAT" value="FITS" array="1">
  <PARAM datatype="char" name="INPUT:imagetype" value="best" array="1">
  <PARAM datatype="char" name="INPUT:inst" value="acs,wfpc2,nicmos">
  <PARAM datatype="int" name="INPUT:hrcmatch" value="0"></PARAM>
  <PARAM datatype="double" name="INPUT:zoom" value="1.000000">
  <PARAM datatype="double" name="INPUT:autoscale" value="99.500000">
  <PARAM datatype="int" name="INPUT:asinh" value="1"></PARAM>
  <PARAM datatype="char" arraysize="*" name="refframe" ucd="VO">
  <PARAM datatype="char" arraysize="*" name="projection" ucd="VO">
</TABLE>
```

User interfaces

Search, browse, plot, explore

Browser-based interfaces

fits2web image viewer (MOS_597 ACS/WFC F775W/F625W (color) CL0152-1357-POS1)

MOS_597 ACS/WFC F850LP/F775W/F625W (color) CL0152-1357-POS1

Lighter Darker Invert advanced contrast controls

☐ DAOphot ☐ SExtractor ☐ SDSS ☐ 2MASS ☐ GSC2 ☐ FIRST ☐ GALEX

Type 'V' for line plot
Type 'C' for column plot
Type 'S' to search

5705 4898
01:52:42.200 -13:57:35.77
0.001329584

based on fits2web

- Services

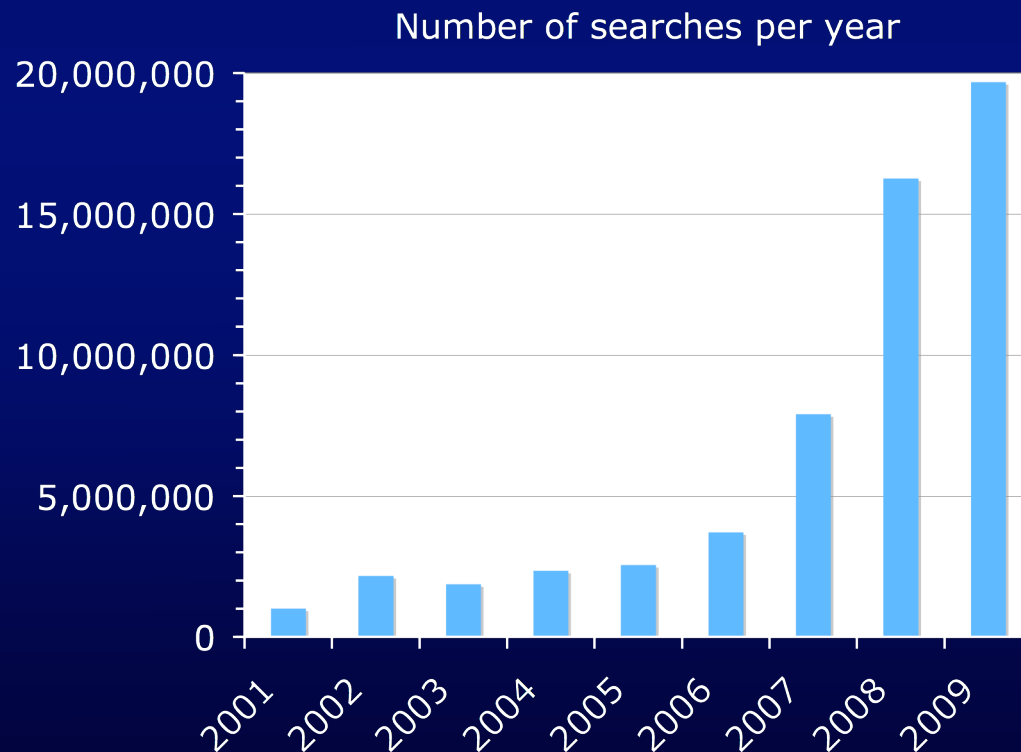
VO services, data retrieval, image cutouts, ...

(UIs are built around VO services)

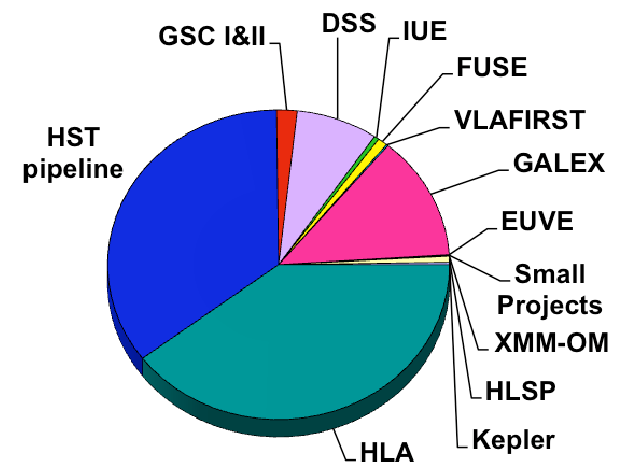
The HST/MAST archive: 1 minute summary



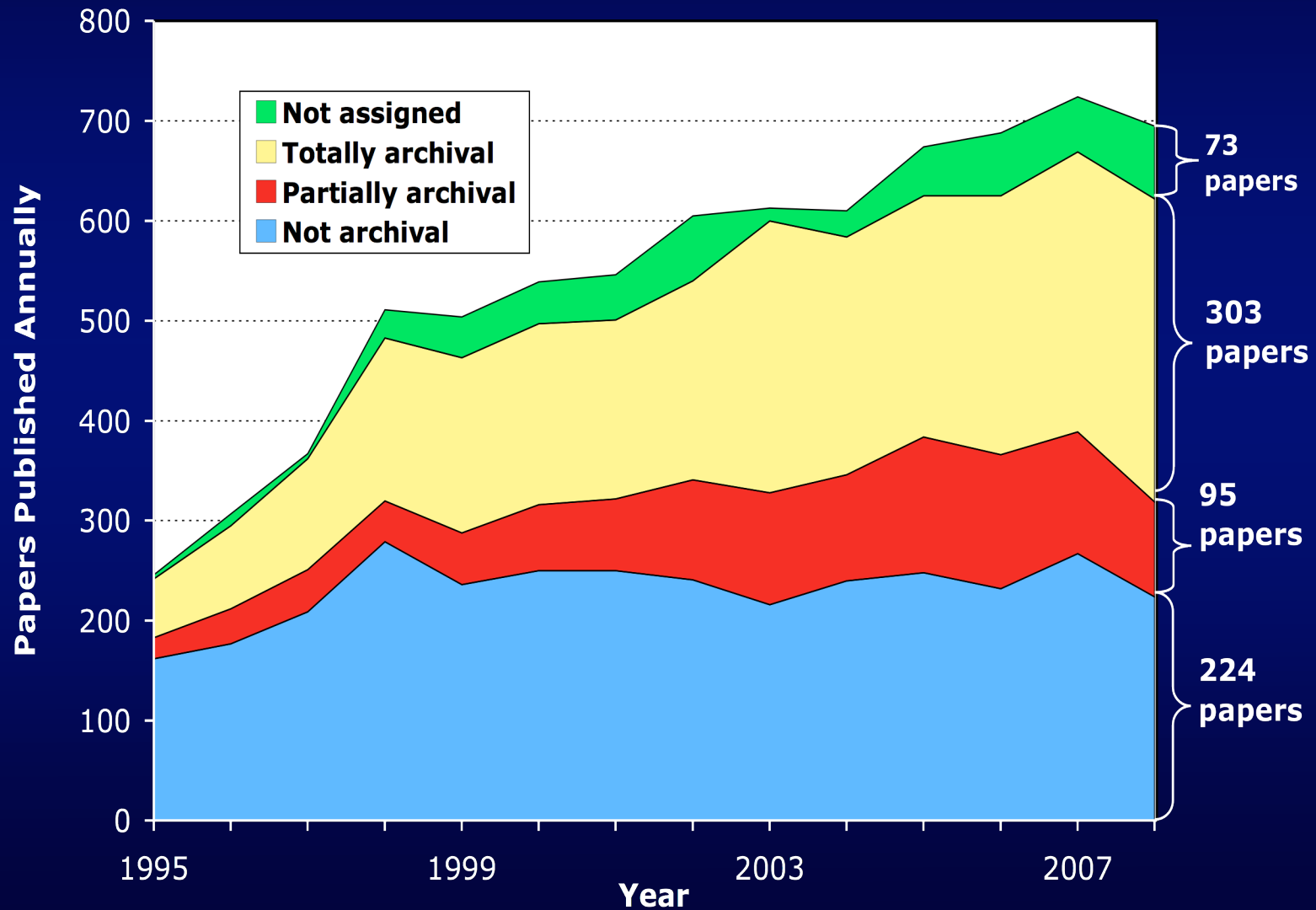
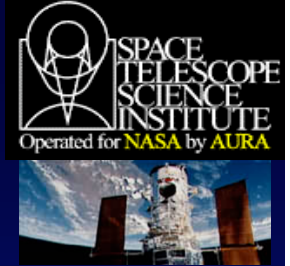
- > 125 TBytes (43 TB HST, 42 TB HLA)
- Ingest rate: 7 TB/yr
- Retrievals: 48 TB/yr
 - Distributed volume ~7x ingest



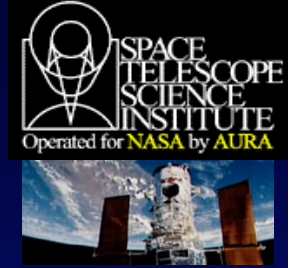
- Data holdings



HST publication statistics

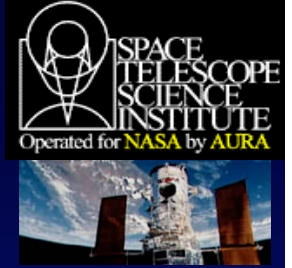


The science impact of data archives



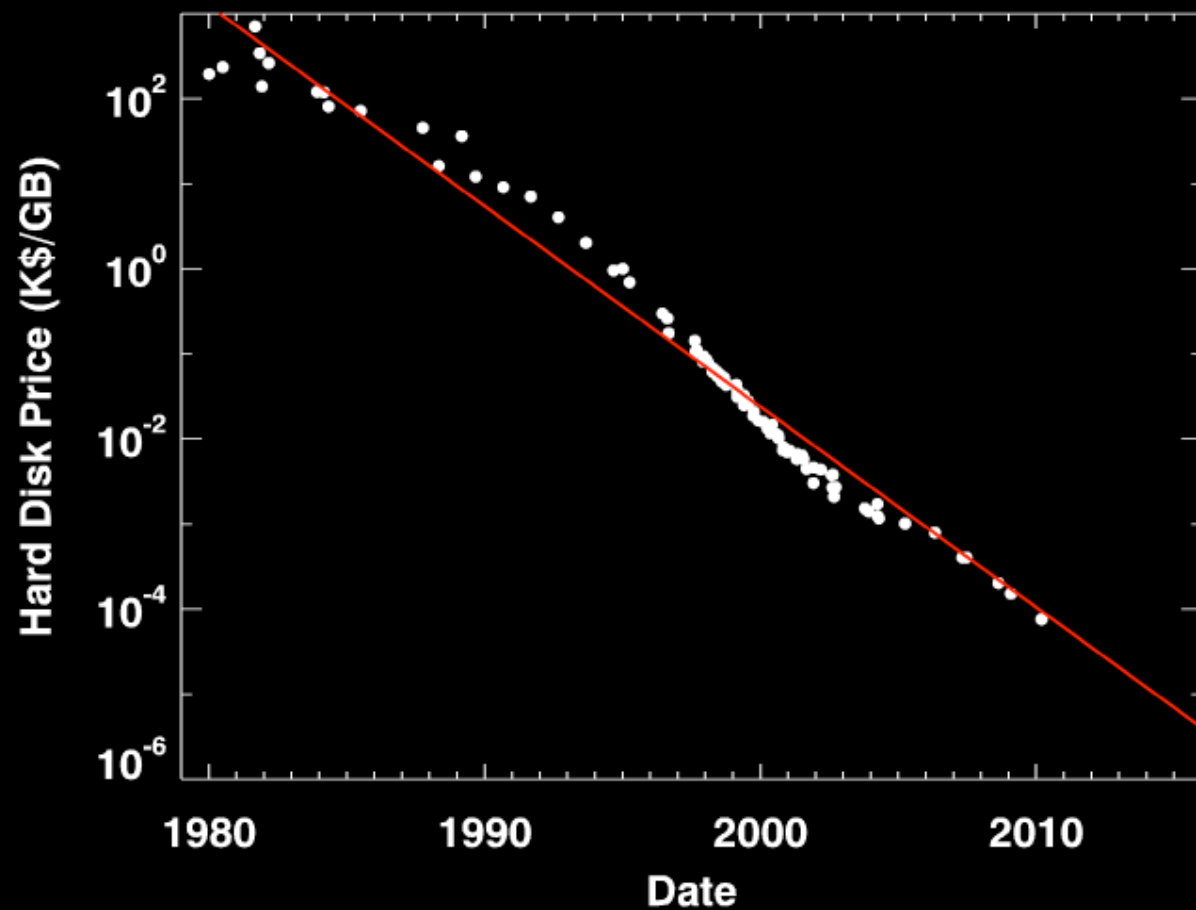
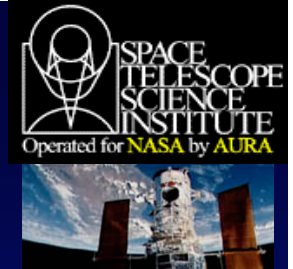
- In the end, the Hubble mission's science impact will certainly be dominated by archival papers
 - The unique value of HST data ensures a long lifetime
 - Increasing the usage of the archive has a direct, measurable impact on science
- See our decadal survey paper for more details:
 - “The High Impact of Astronomical Data Archives”, White et al (2009astro2010P..64W)
 - Archival papers are cited at the same rate as PI-led papers
 - They are also represented as expected among the most highly cited papers
 - Other missions (e.g., Chandra) show similar statistics

History of the Hubble Archive

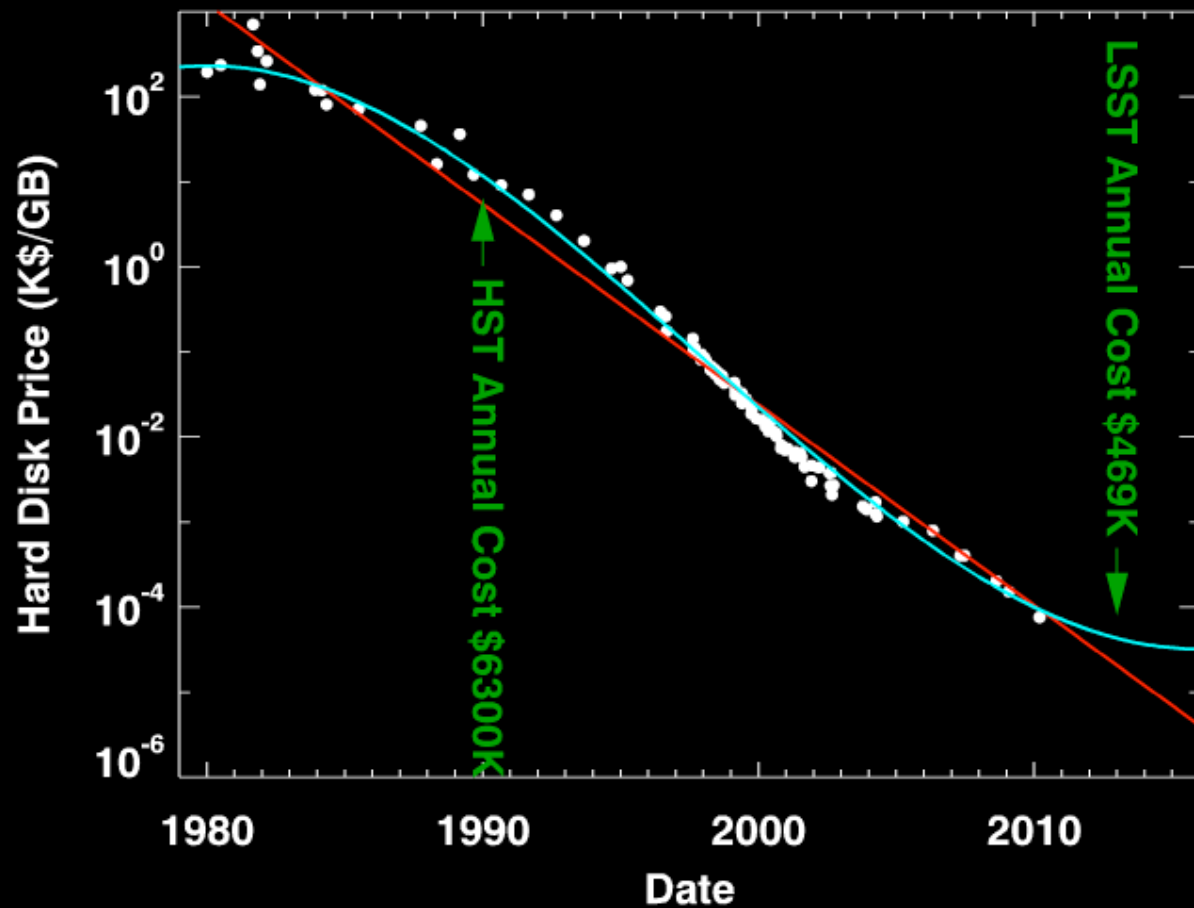
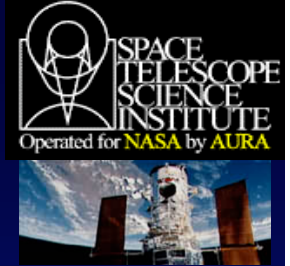


- Operation began at HST launch in April 1990
 - Set new standards for public data access, standard calibrated products, etc.
- Many generations of data delivery media
 - 9-track tape, Exabyte tape, CD-ROM, DVD, Internet
 - Currently almost all data delivered via network (some large deliveries use disk drives)
- Many generations of data storage
 - Various optical and magneto-optical disks, jukeboxes, ...
 - Currently safestore is on UDO
 - Working archive is on spinning disk

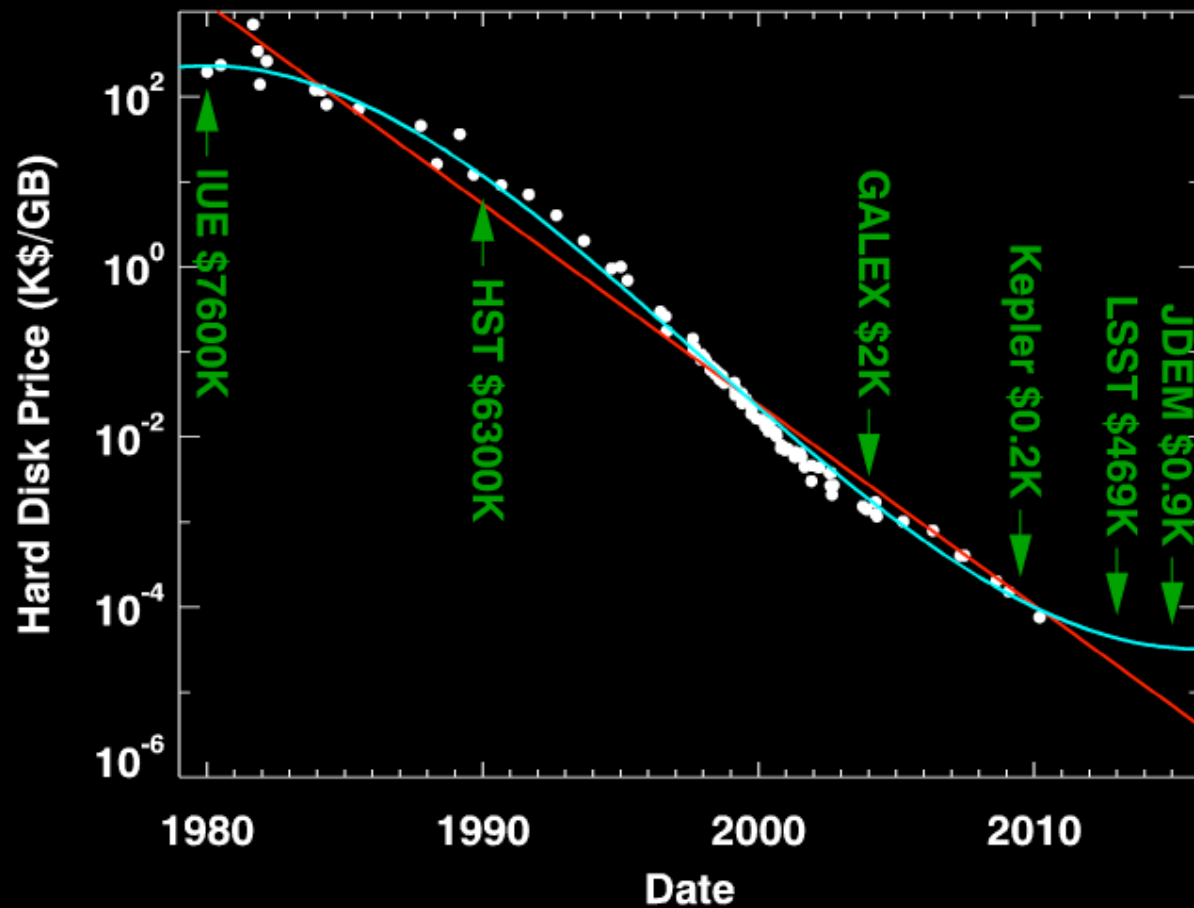
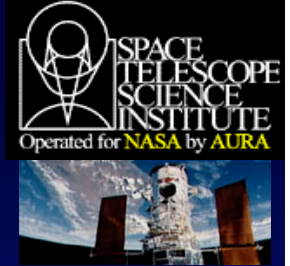
Disk cost per Gigabyte



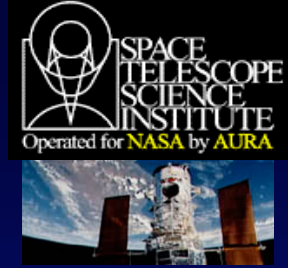
Disk cost per Gigabyte



Disk cost per Gigabyte

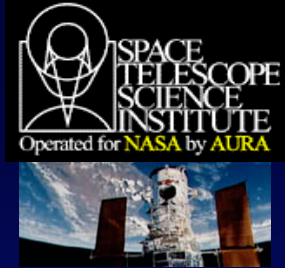


Progress in the archives



- Archives at STScI tend to advance in a leapfrog fashion due to rapidly evolving computer capabilities
 - New capabilities in web interfaces have a similar pattern
- We are now moving into a phase of consolidation and unification
 - Unify underlying databases and server architecture
 - Merge best features into common user interfaces and (VO) web services
 - Created Archive Team to bring together people from all different missions and projects

Selected recent activities



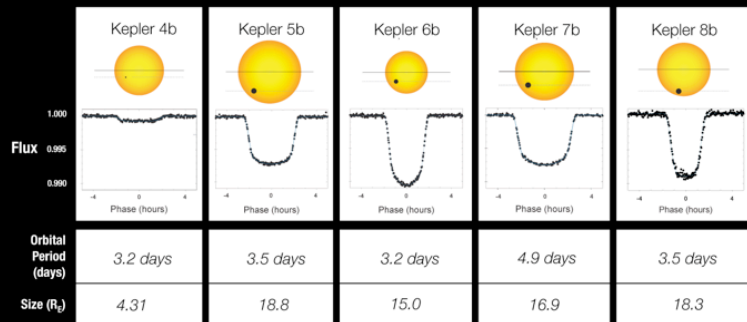
- Starview on the web
 - <http://starview.stsci.edu/web/>
 - Replaces the old Java Starview application
 - Provides user-customizable access to all fields in HST databases
 - Implemented within web browser

A screenshot of the StarView web application running in a Mozilla Firefox browser. The browser's address bar shows "http://starview.stsci.edu - Mozilla Firefox". The application interface includes a "Screens" sidebar with a tree view of instruments (ACS, ALL, COS, FGS, FOC, FOS, FUSE, HRS, HSP, NICMOS, STIS, TV1, TVL, WFC3, WFPC, WFPC2). The "COS" instrument is selected, showing a list of "COS Best Reference Files". The main panel displays a "Search" form with fields for Dataset Name, Proposal ID, PI (last name), RA, Dec, Start Time, Release Date, Target Name, DETECTOR, OPT_ELEM, APERTURE, CENWAVE, OBSTYPE, OBSMODE, and csr_best_flatfile. Each field has a checkbox to the right, indicating whether it is included in the search. The "Search" button is at the top right of the form, and a "Cancel" button is next to it. A "SQL" button is also visible. The interface is user-customizable, allowing users to select which fields to include in their searches.

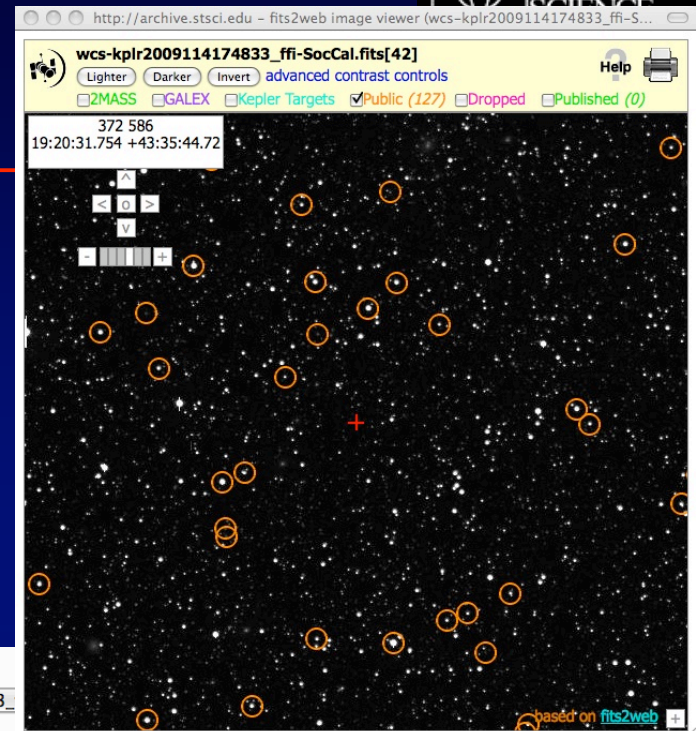
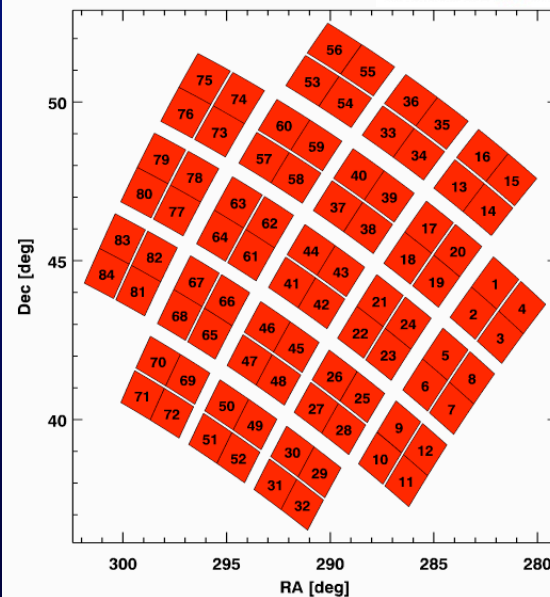
Selected recent activities

- Kepler
 - New mission with significant public data available now
 - >10,000 public light curves
 - Using HLA image display tool

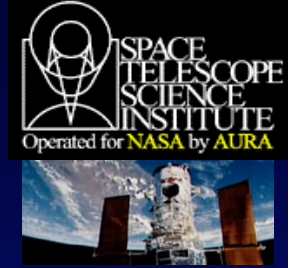
Transit Light Curves



Display channel: 42
for FFI image: wcs-kplr2009114174833



Hubble Legacy Archive: DR4



- HLA DR4 release on March 24, 2010

- DR3: May 2009, DR2.5 (Dec 2008), DR1: Feb 2008
- <http://hla.stsci.edu>

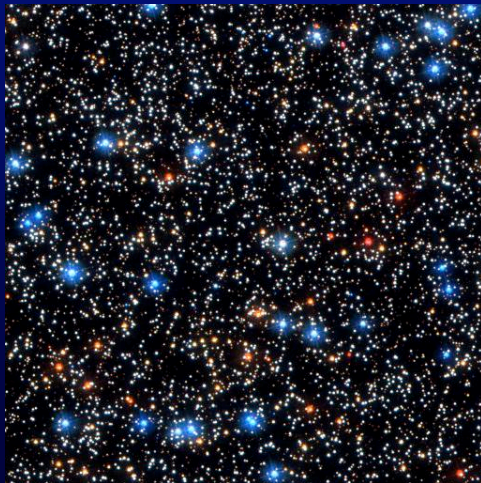
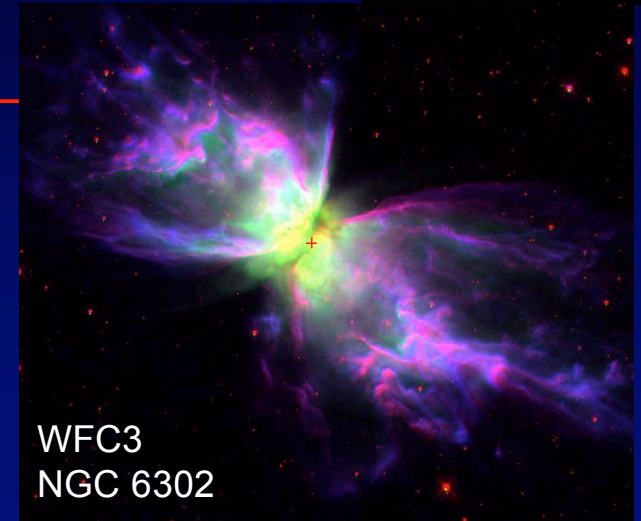
- New data products:

- NICMOS: now 89% complete
- WFPC2: reprocessed, 96% complete
- ACS: 99% complete (includes data public by 2-24-2010)
- WFC3: early release science data
- New high-level science products (3CR snapshots, Andromeda, Carina Nebula, GOODS NICMOS, ANGRRR, SM4 ERO)

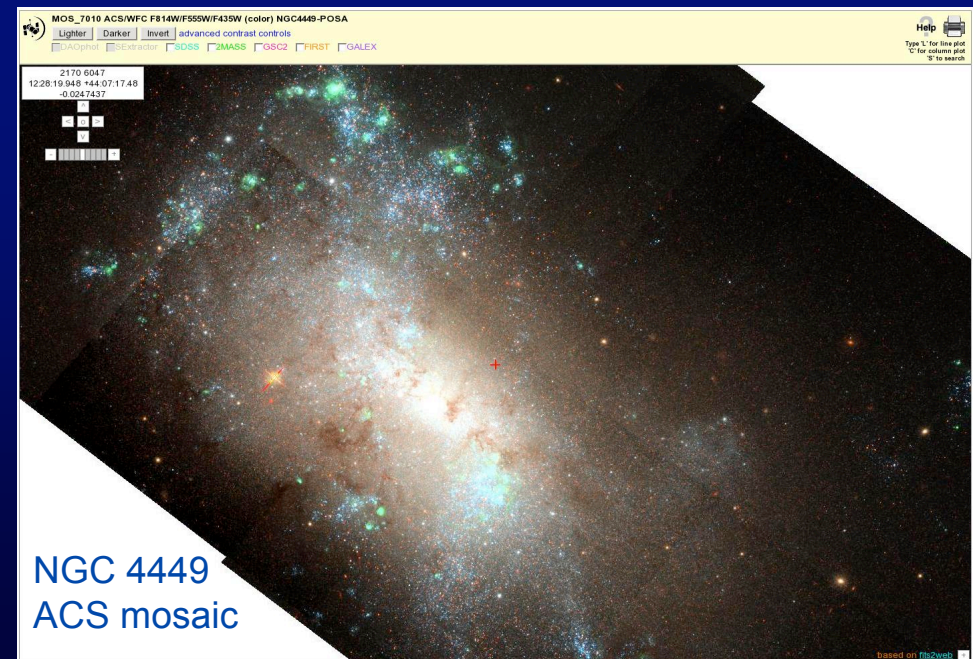


Hubble Legacy Archive: DR4

- User interface enhancements:
 - Dynamic plotting tool for spectra & image cuts
 - All instruments (including COS, WFC3, proprietary data) included in search and footprints view
 - Consolidated, searchable Help Center
 - Interactive display shows pixel value at cursor
 - ... and much more



Omega Cen
WFC3
F336W/F275W/F225W



Near-term Plans: Multi-mission



- Infrastructure redesign
 - Moving all databases to SQL Server
 - Common architecture for Linux/Windows server clusters with failover for high reliability
 - Use virtual machines for maximum flexibility
 - Network bandwidth upgrade
- MAST/HLA consolidation activities
 - Incorporate all MAST missions in HLA footprint database
 - Enable searches for all missions through HLA-style interface
 - Develop new archive portal
 - Unified search interface, “iGoogle-style” customization
 - Unified metrics for tracking usage and performance
- User interface enhancements
 - Science keyword vocabulary (advanced literature index to data)
 - New footprints interface
 - Moving target search

Near-term Plans: Hubble Legacy Archive



- Post-DR4 data processing (complete soon)
 - NICMOS SAAClean, WFPC2 source list processing
- New data structure for FITS images
 - Additional data layers (with compression?)
 - More accurate header keywords
- New & better data products
 - WFC3 enhanced images
 - NICMOS source lists
 - Sky equalization algorithm
 - Improved absolute astrometry (covering all exposures)
 - Relax mosaic restrictions (wider time intervals, handle more images)
 - Mosaics for WFPC2, WFC3, NICMOS
- New user interface tools
 - Footprint interface
 - Better tools for spectra
 - Integrate with other missions (MAST, Chandra, Spitzer)



Near-term Plans: The End of On-the-Fly Reprocessing (OTFR)



- HST data are reprocessed from scratch when retrieved from the archive
 - Uses raw spacecraft plus latest calibration data and pipelines
 - Advantages of OTFR:
 - Best calibration during period when instrument modeling is improving rapidly
 - Reduces storage since raw data is smaller than calibrated data
 - Disadvantages of OTFR:
 - Introduces significant delays in delivering data (~ 0.5–1 hour)
 - Limits archive interface capabilities (e.g., image cutouts)
 - Greatly complicates archive mirroring
- Plan: Maintain a cache of calibrated data and carry out recalibration offline (in the background)
 - Storage and processing are cheap
 - Will also facilitate merger of data collections (HST pipeline, HLA, community-contributed HLSP, ...)
 - Approach pioneered by ST-ECF and CADC

Near-term Plans: Related Activities



- MAST Senior Review: Spring 2011
 - Schedule for new capabilities linked
- ST-ECF closing: end of 2010
 - ST-ECF has provided:
 - archive mirror site
 - advanced prototyping (e.g., original on-the-fly calibration)
 - valuable instrumental expertise (e.g., grism extraction software)
 - Working on migrating data and interface capabilities
- Establish new European mirror site at ESAC
 - Will operate through the data cache (no OTFR pipelines)
 - Unlikely to replace ST-ECF instrumental expertise