The STScI Archive and Data Center

Overview of current and planned activities
Large and treasury proposal plans

Rick White

STUC

2010 November 1
The STScI Archive and Data Center

- Overview of the STScI archive and missions
- Selected recent activities
- Near-term plans
- Large & treasury programs (including MCTP)
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, …)
  - Separately funded by NASA HQ
  - 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**
  - ~$10^6$ HST observations (plus other missions)
  - Documentation, publication links, …

**User interfaces**

Search, browse, plot, explore

**Browser-based interfaces**

- **Services**
  - VO services, data retrieval, image cutouts, …
  - (UIs are built around VO services)
The HST/MAST archive: 
1 minute summary

- > 130 TBytes (50 TB HST, 42 TB HLA)
- Ingest rate: 7 TB/yr
- Retrievals: 49 TB/yr
  - Distributed volume ~7x ingest

Data holdings

- HST
- GSC I&II
- DSS
- IUE
- FUSE
- VLA-FIRST
- GALEX
- Small Projects
- XMM-OM
- HLSP
- Kepler

Distributed data volume

Graph showing the increase in millions of searches per year from 2001 to 2009.

Graph showing the distribution of data volume across different projects from 0.1 TBytes to 100 TBytes.
Selected recent activities

- **Infrastructure upgrades**
  - New SQL Server database systems (ops, test, dev)
  - New compute servers
  - Increased disk storage
- **Internet connection upgrade in progress**
  - January 2011: Lifecast will be new ISP
  - Bandwidth increase from 100 to 500 Mbit/sec
- **MAST Users Group met in July 2010**
  - Report is online: [http://archive.stsci.edu/mug/](http://archive.stsci.edu/mug/)
Selected recent activities

- **Kepler**
  - Quarters 0/1 public data release on June 15: >216,000 public light curves!
  - 590,000 searches and 4.4 TB of data distributed through June!
New High Level Science Products

HLA DR4 release on March 24, 2010
  - http://hla.stsci.edu

New data products:
- NICMOS (89% complete)
- WFPC2 (96% complete)
- ACS (99% complete for 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)
  - July 2010: ~ 48,000 ACS grism spectra (from ST-ECF)

Many new interface capabilities too

HLA Image of the Month with tutorials
  - E.g., http://hla.stsci.edu/iotm/number5_uranus.html
Near-term Plans: Hubble Legacy Archive

- DR5 release planned for January 2011
- New data products
  - WFC3 enhanced images
  - COS combined spectra (possibly)
  - More and better ACS mosaics
  - Incremental data processing with products from newly public data generated and released automatically
- New user interface tools
  - New footprint interface (more interactive, better integration with table filtering)
  - Various other interface improvements
Near-term Plans: HST/Multi-mission

- Eliminate on-the-fly reprocessing for HST data
  - Cache will store current processed data
- Unify mission data
  - Storage broker
  - Data model
  - Footprints for all missions
- Unify user interfaces:
  Archive portal
  - Integrate best of MAST, HLA, GALEX interfaces

NGC 4449
ACS mosaic
Near-term Plans: Multi-mission

• Cross-archive center activities
  – Chandra footprint database
  – ADS/Chandra/MAST data integration
  – Great Observatories archive user workshop (at STScI)
  – Develop college (undergrad/grad) course materials on archive use

• Preparing for MAST Senior Review at NASA HQ
  – Proposal due April 2011, review May 2011
  – Accelerated schedule
    • 3 years since previous review

WFC3
NGC 6302
Large & treasury programs

• Large & treasury programs generate valuable archival data products
  – Many are nominally required to deliver products to MAST
  – HLSP are heavily utilized
    • Downloaded 10x as often as standard products

• MAST tracks publications and high-level science products by HST program
  – See http://archive.stsci.edu/hst/tall.html for active page with latest information
Multi-cycle treasury programs

- MAST will archive MCTP data products
  - Note data are non-proprietary
  - Teams are generating higher-level products
- Plans are well along for all 3 MCT projects
  - CANDELS
    - Developing team site that will become public as data are released
    - First release expected around the end of the year
    - Built around HLA tools plus FTP access
  - CLASH
    - Plans similar to CANDELS: first release May, releases by cluster
  - PHAT
    - FTP data already available
    - First 4 “bricks” in April, releases every 6 months; higher-level products (stellar parameters, extinction maps, etc.) to follow
Large & treasury program plans

- We would welcome changes to put more “teeth” into requirements for data product delivery
  - E.g., could include reports for previous projects in materials given to the TAC
  - MAST tracks this info already and could easily generate reports by proposal ID
  - This information is already available on the TALL page: http://archive.stsci.edu/hst/tall.html