



Hubble Legacy Archive (HLA) Status Update

Stefano Casertano

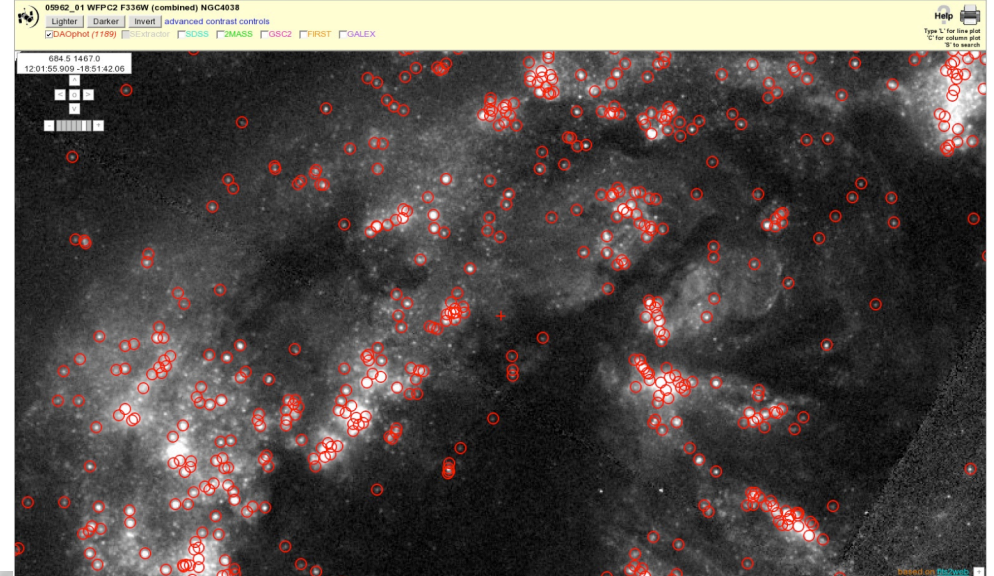
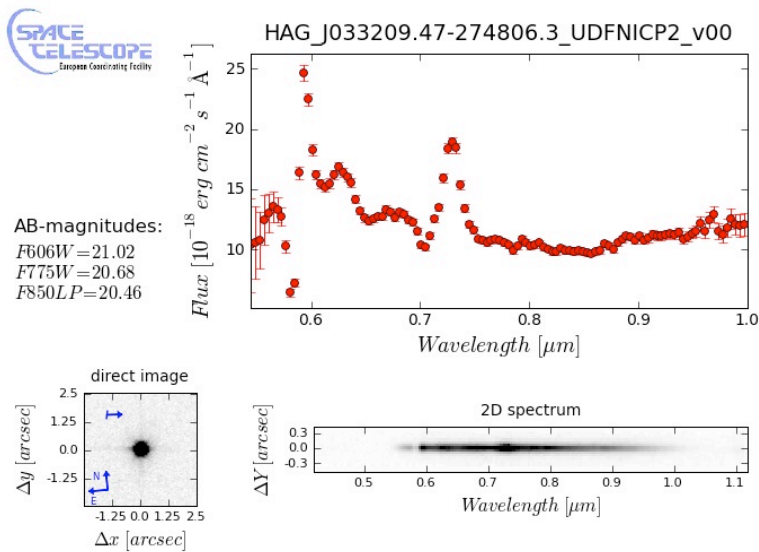
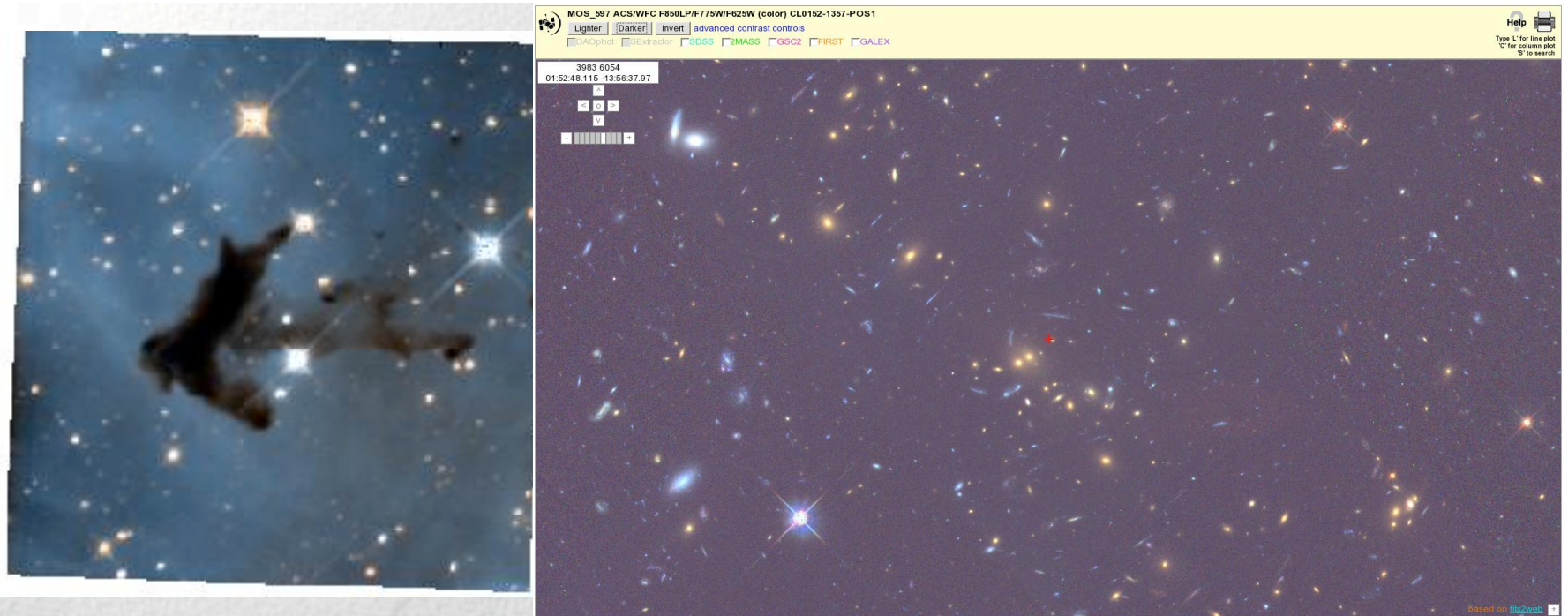
STUC

November 12, 2009



Recent news and plans

- Successful Data Release 3 (May 2009)
 - New classes of data
 - NICMOS combined images
 - ACS extracted GRISM spectra
 - WFPC2 source lists
 - Prototype ACS mosaics
 - Interface upgrades
 - Improved table interaction across views
 - Thorough revision of help information
 - Follow-up work needed
 - More mosaics
 - Completion of NICMOS, ACS GRISM
 - Improvements in WFPC2 and ACS source lists
- Extensive work on upgrading the HLA infrastructure
- Next release (DR4) planned for January 26, 2010





Infrastructure Upgrades

- Since DR3, our focus has been on improvements in our infrastructure
 - Automated source list processing and ingest (done)
 - New organization for help information (done)
 - Completion of footprint database (almost done)
 - Automated data processing pipelines (ongoing)
 - Data validation (ongoing)
 - Preparation of paper for publication (ongoing)
 - Changes to data formats (discussion ongoing)
 - Modular pipelines with code shared across instruments (just started)
 - Will include automated quality verification steps and reports
 - Improved algorithms for sky equalization, image registration, noise estimate under development



Automated pipeline control

- Replace manually-run pipelines with scripts started by trigger files using the NHPPS control system
 - Provide easier control, more stable performance and reduced need for manual intervention
 - First step in building an integrated automated processing system
 - Built using current pipelines; upgrades in the plans for post-DR4 work
 - Currently require manual monitoring and quality verification; some default quality controls are planned
- Current status
 - ACS pipeline ported and tested; data validation steps included
 - New NICMOS pipeline built by modification of ACS system
 - ACS mosaic pipeline built from adapted ACS scripts
 - Source list processing pipelines developed
 - Unification of scripts across instruments in progress
 - Mosaic and NICMOS pipelines already used for DR3 (not in NHPPS version); other pipelines will be inaugurated for DR4



Streamlined source list ingest

- The source lists database currently drives the ability to overplot sources on HLA-processed ACS and WFPC2 image data
- Source list ingest was a manual, time-consuming operation through DR3
- We have modified the ingest scripts to run automatically as part of the processing pipeline
- In the process, we have reviewed the formats of the source lists and adjusted their content for consistency across instruments and versions
- We plan to develop tools to allow users to generate desired source lists on the fly from the database (currently lists are only available in pregenerated, fixed-format files)
- Source list availability information will also be accessible to the cone search; the interface will be updated to allow users to include source list properties as search criteria



Completeness of footprint database

- In preparation for APT support, the database of HLA-processed data is going through a full review
 - All HST data will be identified with an HLA processing status flag
 - In most cases, the lack of HLA processing is deliberate and due to one of several causes:
 - Lack of scientific content (e.g., Earth flats)
 - Availability of high-level products (e.g., COSMOS)
 - Data quality problems
 - Inability to properly combine images (planetary data, NICMOS focus runs)
 - Some cases are not fully understood, but all will be documented
 - When HLA-processed data are not available, the footprint database will be populated from the science data table
- The full database is nearly complete, and will be used extensively to ensure the desired completeness for HLA processing



New data formats and content

- HLA data formats defined during early development period
 - Largely follow Multidrizzle conventions with some modification
 - Provide much of the information users need
 - However:
 - Some information not propagated correctly
 - Noise model incomplete
 - Headers store information in cumbersome, hard-to-retrieve way
- Agreement between the HLA partners (STScI, CADC, ST-ECF) (March 2009) to revisit data formats
- General agreement on overall desired properties
- Proposed new format now under discussion
 - Input needed!



Proposed new format

- Multiextension FITS file (as now)
- Sky equalization instead of subtraction
 - Match sky background in overlapping images and subtract the difference
 - Current sky subtraction fails for extended/diffuse sources
- Three data layers
 - Image data in e/s (COUNTRATE)
 - Inverse variance from subtracted data (IVM)
 - Effective exposure time (EXP)
 - Allows full noise model:
 - $\text{NOISE}^2 = 1/\text{IVM}^2 + \text{COUNTRATE}/\text{EXP}$
 - To avoid bias, COUNTRATE should be from a data model, not directly from the data values
- One or more integer layers
 - Number of exposures with/without rejection
 - May add flag image to propagate flags in input images
- Context image probably removed



```
hst_10520_01_wfpc2_total_wf_drz.fits
File Edit
FILENAME= 'hst_10520_01_wfpc2_f555w_wf_drz.fits'
ASN_MTYPE= 'PROD-DTH'
NDRIZZIM = 16 / Drizzle, No. images drizzled onto output
D001VER = 'Callable DRIZZLE Version 0.7 (4th Apr 2005)' / Drizzle, task version
D001GEOM= 'User parameters' / Drizzle, source of geometric information
D001DATA= 'u9gz0101m_c0h.fits[sci,1]' / Drizzle, input data image
D001DEXP= 1600.0 / Drizzle, input image exposure time (s)
D001OUDA= 'hst_10520_01_wfpc2_f555w_wf_drz_sci.fits' / Drizzle, output data imag
D001OUWE= 'hst_10520_01_wfpc2_f555w_wf_drz_weight.fits' / Drizzle, output weight
D001OUCO= 'hst_10520_01_wfpc2_f555w_wf_drz_context.fits' / Drizzle, output conte
D001MASK= 'u9gz0101m_c1h_final_mask1.fits' / Drizzle, input weighting image
D001WTSC= 110028.8821198162 / Drizzle, weighting factor for input image
D001KERN= 'square' / Drizzle, form of weight distribution kernel
D001PIXF= 1.0 / Drizzle, linear size of drop
D001COEF= 'u9gz0101m_c0h_coeffs1.dat' / Drizzle, coefficients file name
D001XGIM= '' / Drizzle, X distortion image name
D001YGIM= '' / Drizzle, Y distortion image name
D001LAM = 544.2935 / Drizzle, wavelength applied for transformation
D001SCAL= 2.196257796340602 / Drizzle, scale (pixel size) of output image
D001ROT = -1.15224729 / Drizzle, rotation angle, degrees anticlockwise
D001XSH = -5.26910851998025 / Drizzle, X shift applied
D001YSH = -21.4485678167905 / Drizzle, Y shift applied
D001SFTU= 'output' / Drizzle, units used for shifts
D001SFTF= 'output' / Drizzle, frame in which shifts were applied
D001EXKY= 'EXPTIME' / Drizzle, exposure keyword name in input image
D001INUN= 'counts' / Drizzle, units of input image - counts or cps
D001OUUN= 'cps' / Drizzle, units of output image - counts or cps
D001FVAL= 'INDEF' / Drizzle, fill value for zero weight output pix
D001INXC= 400.5 / Drizzle, reference center of input image (X)
D001INYC= 400.5 / Drizzle, reference center of input image (Y)
D001OUXC= 775.5 / Drizzle, reference center of output image (X)
D001OUYC= 775.5 / Drizzle, reference center of output image (Y)
D002VER = 'Callable DRIZZLE Version 0.7 (4th Apr 2005)' / Drizzle, task version
D002GEOM= 'User parameters' / Drizzle, source of geometric information
D002DATA= 'u9gz0101m_c0h.fits[sci,2]' / Drizzle, input data image
D002DEXP= 1600.0 / Drizzle, input image exposure time (s)
D002OUDA= 'hst_10520_01_wfpc2_f555w_wf_drz_sci.fits' / Drizzle, output data imag
D002OUWE= 'hst_10520_01_wfpc2_f555w_wf_drz_weight.fits' / Drizzle, output weight
D002OUCO= 'hst_10520_01_wfpc2_f555w_wf_drz_context.fits' / Drizzle, output conte
D002MASK= 'u9gz0101m_c1h_final_mask2.fits' / Drizzle, input weighting image
```




Proposed new format (cont.)

- Rearrange header information
 - Information from individual images stored in FITS table extension
 - Current mechanism involves ad-hoc solution
 - More flexibility, searchability
 - Keyword names can be propagated without change
 - Solves performance issues for images with many components
 - Remove header keywords that are not applicable to combined image
- Modify naming convention to ensure uniformity across instruments
- Adjust astrometry for all products
 - Use matches to external catalogs if available (as now)
 - Use improved guide star positions and focal plane model otherwise



Data verification

- We are going through multiple data verification procedures
 - Cross-instrument, multi-epoch comparisons
 - Comparisons with independent processing of the same data
 - Comparisons to other (non-HST) data
- Most results are positive
 - Photometry, astrometry consistent across many years
 - Uncertainties comparable to expectations
- Identified some bugs in existing code
 - WFPC2 AREA mode images do not run correctly through Multidrizzle
 - Only affects 300 images ($<0.3\%$)
 - Corrected code will be released soon
 - Weights are not computed correctly for WFPC2 images
 - Led to fewer than expected sources found by SExtractor at time of DR3
 - New weight calculation available for DR4
 - Improved source lists will be released
 - Zero points, aperture corrections not completely consistent
 - Different apertures used for different instruments
 - Revisit all zero points and aperture corrections for uniformity
- Many tests will be automated and included in the new pipelines

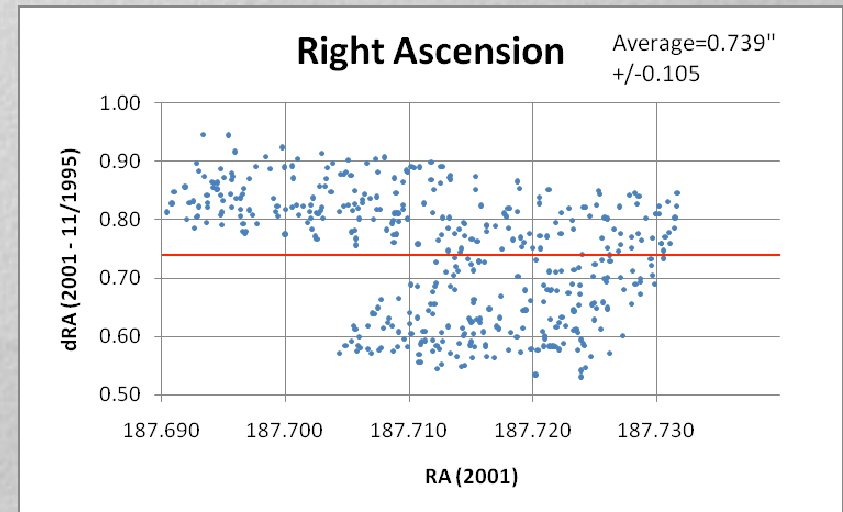
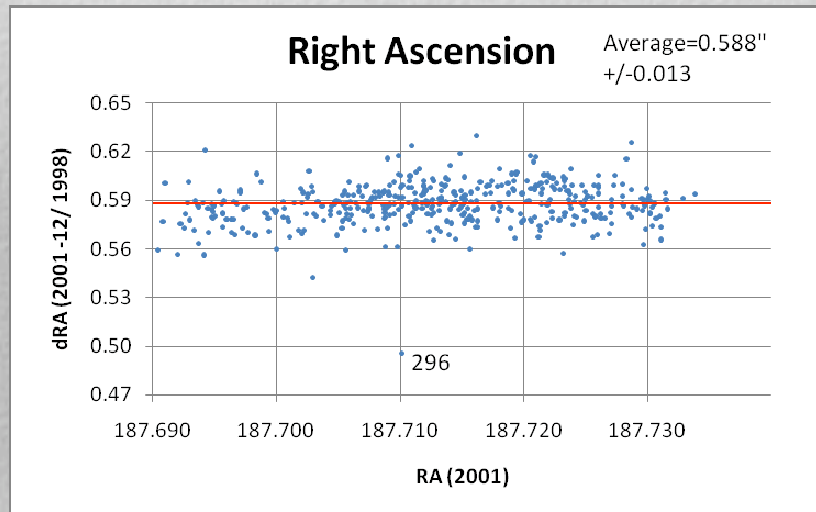
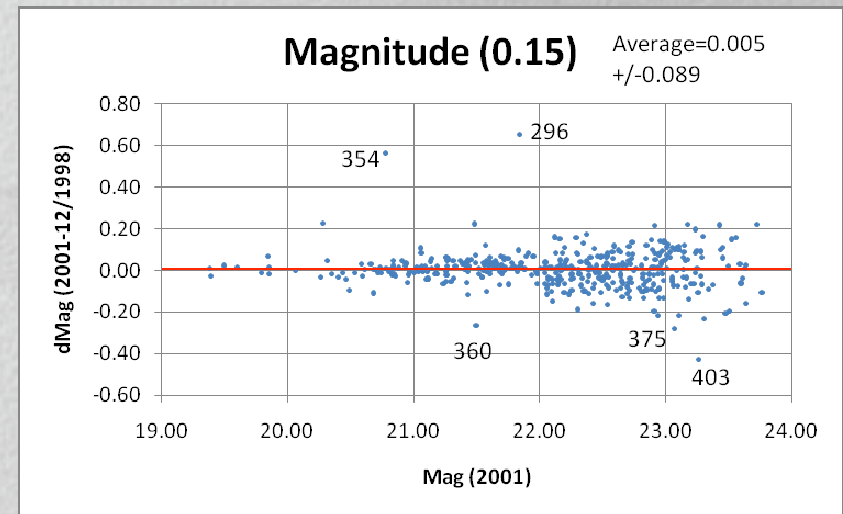


Comparison between WFPC2 observations over 6 years

Magnitudes consistent within errors

Most astrometry consistent except for systematic shift (expected)

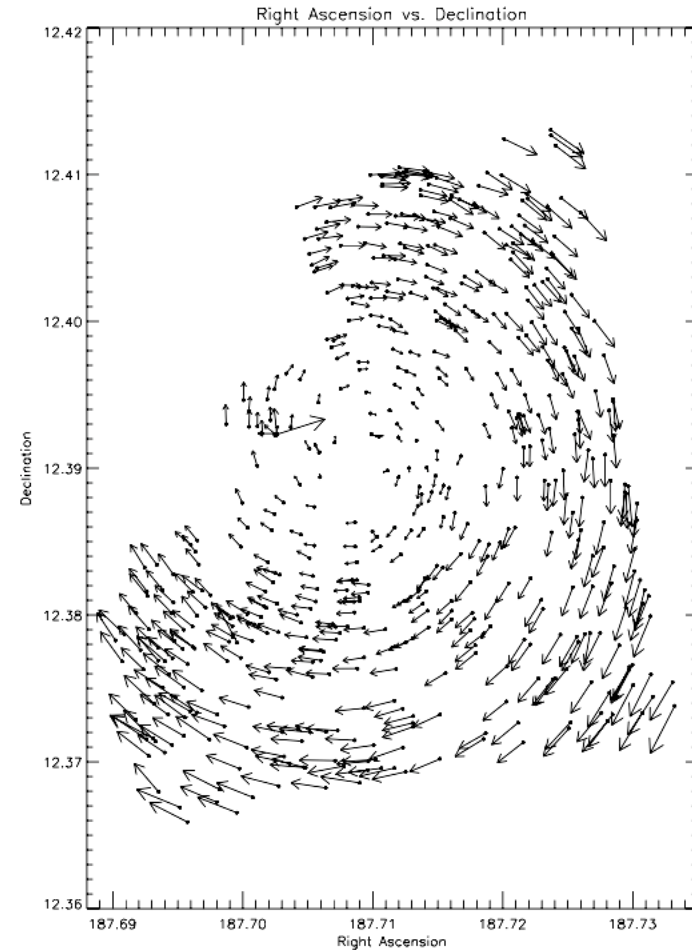
Some unusual residuals seen





WFPC2 comparisons (cont)

- Unusual shifts turn out to be pure rotation
- Rotation angle 0.13 deg
- Problem was known and described in WFPC2 documentation





Coordination with other Centers

- Code base shared with CXC
 - September meeting with AHLA, JHU, and CXC to exchange footprint service software and database schema
- CXC has successfully configured an HLA VO data access layer service for use with Chandra footprints
- Goal is ultimately to share footprint data across missions and enable coordinated searches
 - Collaboration promising for sharing VO compliant data services
 - New footprint display under development
 - Additional interface features will be needed to make best use of these capabilities
- Interaction with SSC progressing slowly



Plans for DR4

- Data Release 4 is planned for January 2010
 - A test version will be available at the AAS
 - Testing will most likely be completed after the AAS for a release around January 26
- Data slated for inclusion in DR4 are:
 - COS and WFC3 data as high-level science products (as released by the Institute and the Science Teams)
 - Independent HLA processing of COS and WFC3 data is not planned until the following release
 - New NICMOS combined images for all reprocessed data (~90%)
 - New WFPC2 combined images for most reprocessed data
 - Processing time may prevent a complete WFPC2 release
 - Remaining data will be made available through incremental releases
 - Several new ACS mosaics (~50)
 - Only mosaics that process correctly though the available pipeline will be included
 - Pipeline upgrade is not planned until the following release
 - Revised WFPC2 source lists
 - Processed data for recent observations



Plans for DR4 (cont)

- Interface and infrastructure upgrades for DR4 include:
 - Full support for APT through the footprint database
 - A new Help Center integrating the HLA user information
 - Inclusion of source list information among user-searchable parameters
 - Upgraded plotting package for spectra, image cuts
 - Possibility to deliver source lists directly from the database
 - User-customizable options to be added in future releases
 - Prototype of new footprint visualization tool (for AAS demo, not included in DR4 release)



Long-term plans

- High-priority items post-DR4 include:
 - Implementation of new data formats and structure
 - Automated footprint update
 - New or improved pipelines for mosaics, new instruments
 - Multi-visit mosaics for NICMOS, WFPC2, WFC3
 - Automated processing for most instruments
 - Interface to spectral container for one-dimensional spectral data
 - Inclusion of Chandra and Spitzer footprints
 - New footprint interface