DRAFT - Plans for the Development of a Hubble Source Catalog

A White Paper

April 5, 2012

Brad Whitmore, Stefano Casertano, Anton Koekemoer, Steve Lubow, Rick White

1. INTRODUCTION

As of January 1, 2012, Hubble has taken >10**6 exposures, containing >10**9 individual objects. The diversity of this immense dataset, and the slight imperfections (typically 1 - 2") in the registration of HST images taken at different times, make it difficult for the average observer to extract basic information (magnitude, color, variability) for individual objects without substantial effort. An HST Source Catalog will provide this type of information quickly in a form that is useful, searchable, and easily accessible.

Easily accessible catalogs of astronomical objects are a mainstay of astronomy. A recent example is the Sloan Digital Sky Survey (SDSS) catalog, which is largely responsible for the great success of the SDSS project. One might ask the question, "How much science would have come out of SDSS if each astronomer had to access the data themselves and make their own catalog?"

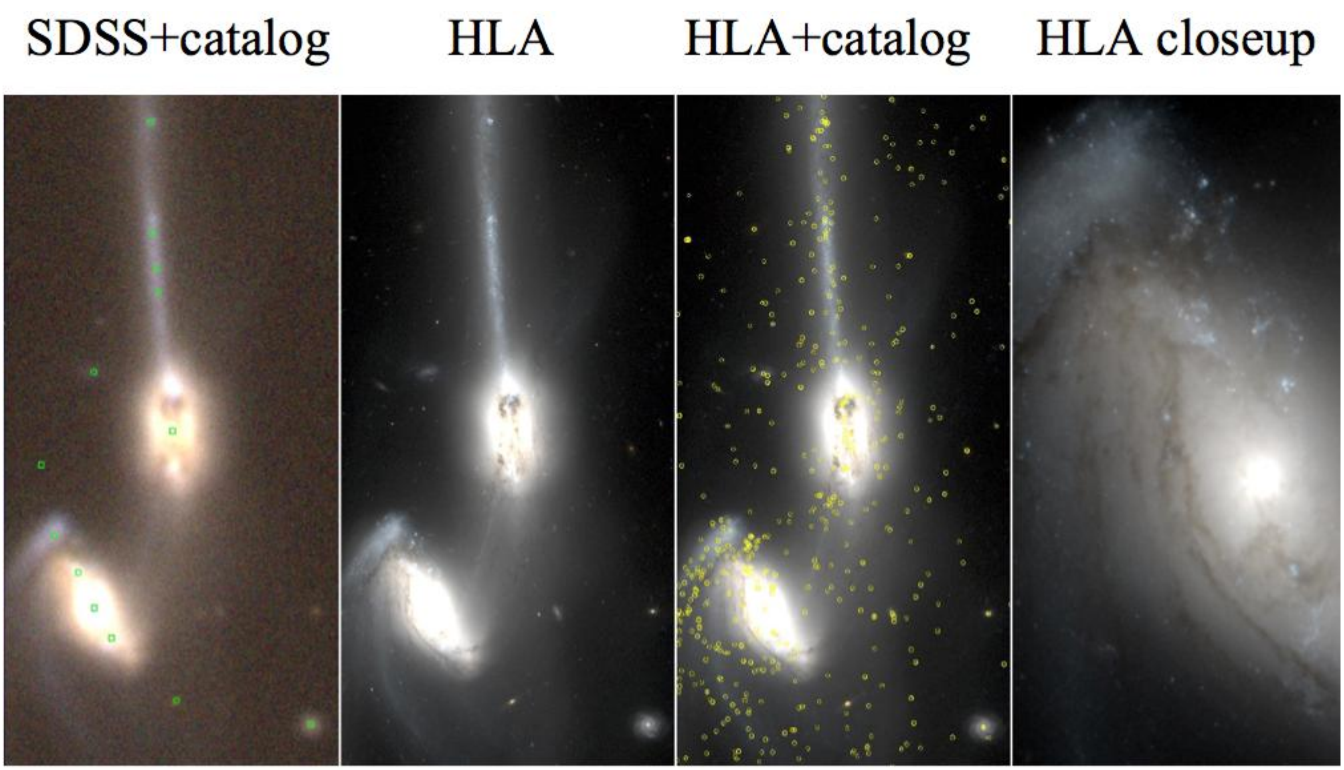

Figure # 1 - SDSS image/catalog with corresponding Hubble image/catalog.

In the near future, sky surveys such as Pan-STARRs and LSST will revolutionize the amount and types of information available for astronomers to mine. However, they will not be able match Hubble's spatial resolution (by a factor of ~10 in angle and ~100 in resolution elements), which provides unique morphological information on sub-kpc scales for galaxies at all redshifts. For example, from the ground most galaxies beyond z~0.5 are unresolved in the large ground-based surveys. Cross-referencing these new surveys with a Hubble Source Catalog (HSC) will leverage HST observations for decades to come. In addition, the ultraviolet and optical components of the HSC will provide an important complement to the IR capabilities of JWST.

2. GOALS and REQUIREMENTS

The overarching goal of the HSC project is to optimize science from the Hubble Space Telescope by developing a single catalog containing the vast majority of all sources ever observed by Hubble.

The baseline goals are to:

B1) - Provide a general, easily accessible, time-resolved catalog of "all" 10 sigma sources observed by the telescope, using the ACS, WFC3 and WFPC2 instruments.

B2) – Cross-reference this catalog with a basic list of existing catalogs (e.g., SDSS, 2MASS, Spitzer, GALAX, Chandra, existing Hubble HLSP, ...) as well as HST spectroscopic datasets that may be relevant.

B3) - Develop tools for easy access and quick-look analysis. These would include a relational database search capability (e.g., all Sc galaxies within 10 Mpc), and a quick-look, time-resolved phenomena tool (e.g., to search for periodicity).

The stretch goals are to:

S1) - Extend the catalog to other instruments (e.g., NICMOS, STIS, ...) and to develop more focused, "value added" catalogs addressing specific science topics (e.g., age estimates for compact star clusters for galaxies within 20 Mpc).

S2) – Cross-reference with a more extensive list of catalogs. Develop a spectroscopic pipeline (similar to the imaging pipeline used in the HLA) and cross match the relevant spectroscopic information with the HSC.

B3) - Develop more advanced tools including a three-plane visualization tool (i.e., tracking the same data points in image - plots - catalog space).

These basic goals will guide the development of a set of formal requirements for the program. A timeline will be developed and tracked, with semi-annual reporting to the Hubble Mission Office. Although it may be premature to estimate the overall time frame until more detailed plans are completed, we envision roughly a 2 – 3 year duration for the development phase, followed by a operations phase using a much smaller staffing level.

3. FOCUS

An important question to ask is who will use this catalog and how will they use it. It might, for example, be argued that most of the best Hubble datasets already have catalogs (e.g., HDF, UDF, GOODS, GEMS, AEGIS, COSMOS, CANDELS ...) and these catalogs are of better quality. In addition, many HST images already have general-use, single-visit source catalogs available via the Hubble Legacy Archive (HLA). However, neither of these catalogs provides users with a homogeneous, cross-identified, ***unique*** list of sources associated with a large fraction of all HST data (WFC3, ACS, WFPC2 for the baseline plan + NICMOS, STIS for the stretch plan). The HSC will also be cross matched with several existing catalogs, covering a wide range of wavelengths (e.g., Spitzer, GALAX, CHANDRA, ...), providing substantially more leverage than any of the existing catalogs.

The crux of the effort in building the HSC is the matching of all this information into a single coherent, easily searchable, general-use catalog that will enable a wide range of new studies, including time-series studies. This will leverage the extensive catalog work that has already been done while at the same time extending these techniques to a much broader set of data, hence increasing their overall scientific value. We also plan to look into the feasibility of building specialized, value-added catalogs designed for more specific questions (i.e., the stretch goals). This will help ensure that the general catalog is as useful as possible, and that particular areas of science with the most favorable cost/benefit ratios are addressed in greater detail.

4. BASIC APPROACH

General-use source lists are already generated in the HLA for ***single-visit*** imaging observations (see, e.g., Fig. 1 for an example). Separate DAOPHOT (point and point-like objects) and Sextractor (extended, point, and point-like objects) catalogs have already been produced for much of the data. Improvements to the existing HLA images will be considered before regenerating a new set of source lists (e.g., using smaller pixels to improve resolution when the number of dithered positions warrant it). The algorithms for producing the existing HLA single-visit source lists will be tailored to the development of the HSC, with significant upfront cost savings to the mission.

Additional effort required to produce the HSC includes the registration of single-visit images of the same region of the sky taken at different times; identification and cross matching of all sources (including combining the DAOPHOT and Sextractor catalogs); combination of data to increase S/N; and collection of time-variability information for each source. In addition, it will be necessary to develop a system that automatically updates the HSC regularly as new data enter the archive.

The Instrument Teams at STScI are steadily improving their image processing (e.g., ACS CTE and bias striping correction, WFC3/IR flat field features, ...), and such improvements will be included in the source material used for the HSC. Instrument scientists will be involved (as consultants) in specific steps in the construction of the catalog.

A prototype study (Steve Lubow/STScI, Alex Szalay/JHU, Tamas Budari/JHU) to cross-match the ACS and WFPC2 single-visit HLA source lists is well under way and is producing encouraging results (i.e., median values of 0.08 mag for the difference between cross matched sources; typical matching accuracies of 10 -25 mas for 40 % of the matches – see Figure 2). We will use this prototype as a springboard for other aspects of the HSC project that require cross identification of sources (e.g., combining the DAOPHOT and Sextractor catalogs). A progress report for this prototype study is available on request.

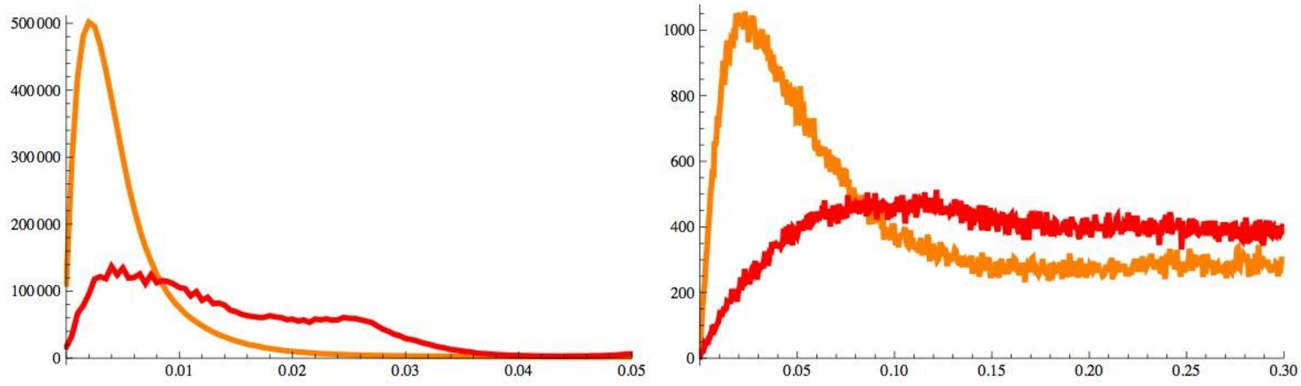

Figure # 2 - Distribution of separations for pairs of matched sources. The horizontal axis is in arc-sec. Red: before astrometric correction; orange: after astrometric correction. Left: A Galactic bulge field of 25 overlapping white-light images and about 8 × 10**5 sources. Right: CDF South UDF field of 194 overlapping white-light images and about 1.4 × 10**4 sources (courtesy of S. Lubow).

As part of the stretch goals, algorithms for the analysis of basic spectroscopic information (e.g., line identification, velocity, equivalent width) will be developed and cross-matched with sources from the HSC.

It is important to note that the HSC will not eliminate the need for astronomers to make more detailed catalogs optimized to address their own specific science goals in many cases. The HSC is designed to be a general-use catalog sufficient for many, but not all users. However, we note that even in the cases where a researcher makes their own specialty catalog, the HSC will still be a useful cross-reference and normalization resource.

5. SOME CHALLENGES

Producing a uniform source list from the wide variety of HST images introduces a number of important challenges that must be addressed and solved. Unlike many survey catalogs (e.g., SDSS), which are based on a uniform set of data (i.e. same filters, same exposure times – although variable PSF and background due to atmospheric conditions), HST observations are taken with a wide range of cameras, detectors, filters, and exposure times.

We will not attempt to include 100 % of all Hubble detections in the HSC. For example, HLA source lists are not produced for some observing modes such as polarization, grism data, and quad filters. Similarly, we would begin with the most popular cameras (ACS, WFPC2 and WFC3) and include others (NICMOS, STIS) as part of the stretch goals, but would not plan to include other instruments where the cost benefit is lower (e.g., WFPC1). These design decisions simplify some of the items discussed below.

1. Uniformity - One of the keys to uniformity is to have reliable uncertainty estimates to use as selection criteria. Our approach will be to use empirically determined “true” errors (e.g., from repeat measurements of the same objects with the same exposure times) to bootstrap our error estimates to be as realistic as possible. Another key is to allow the user to determine the

selection criteria to build their own user-directed catalogs (e.g., only ACS and WFC3 data with F435W, F555W, and F814W observations with exposure times greater than 1000 sec).

2. Completeness - The canonical way to approach this problem is to add artificial sources and determine to what level they are recovered. While this is relatively straightforward for stars on a "blank" background, the problem quickly degenerates into a multi-parameter maze including the size, profile, and shape of an object along with the background level, degree of crowding, and multi-wavelength detection issues (i.e., this is a topic that even very specialized papers do not often deal with). One of the keys is to stay well above the completeness level, hence our goal of using 10 sigma as our detection criteria. In addition, meta-data information relevant to the environment around each source (e.g., background and "local confusion") will be kept, since it can be used to test for dependencies on completeness (e.g., split the catalog into four quartiles based on background and/or crowding and see at what point the results are affected).

3. Use mosaic-based or single-visit-based catalogs ?– One approach to making a master catalog is to add all the images together into a single large, deep mosaic, and determine the sources from the mosaic image. Photometry from each individual image is then measured at these positions. While this approach has several advantages, and indeed is the approach taken in the HLA within a single visit (i.e., the white-light image), practical limitations of making high quality mosaics (i.e., getting all the images perfectly lined up, combining data with different pixels scales such as WFPC2 and WFC3, …) keep this from being a realistic option at present. We therefore plan to start with the single-visit HLA catalogs as our primary building blocks, and consider using a hybrid approach later when adequate mosaics are available.

4. Different spatial resolution - This is especially important when cross matching with external catalogs (i.e., 0.1 arcsec for Hubble – 2 arcsec for Spitzer), but is also an issue when comparing/combining the three primary instrument (WFC3/UVIS - 0.04 arcsec pix, ACS - 0.05 arcsec pix, WFPC2 - 0.10 arcsec pix over most of the FOV, WFC3/IR - 0.13 arcsec pix). For "isolated" point sources, it is possible to use the appropriate aperture corrections to determine total magnitudes, which can then be combined. In crowded regions, we may need to develop software (and make it publicly available) for degrading the resolution of the Hubble image to match the other catalogs.

6. INVOLVEMENT OF THE COMMUNITY

It is important to involve the community in early stages of the HSC project for two basic reasons: 1) to help guide the development of the HSC so that it meets the real, rather than (STScI) perceived needs, 2) to get buy-in from some of the key groups so they will help sell the product.

Five generic use cases have been identified and will help guide the development of the HSC. Small working groups (~6 active researchers in the field) will be established for each topic. In all cases at least one STScI staff member will be a member of the working group and will act as a conduit of communication and information.

The current set of topics being considered are:

1. *Point source photometry* (e.g., color-magnitude diagrams in globular clusters and in the field) [Potential working group members: Tom Brown, Jay Anderson, Jason Kalirai, Peter Stetson,

Abi Saha, Julianne Dalcanton, ...]

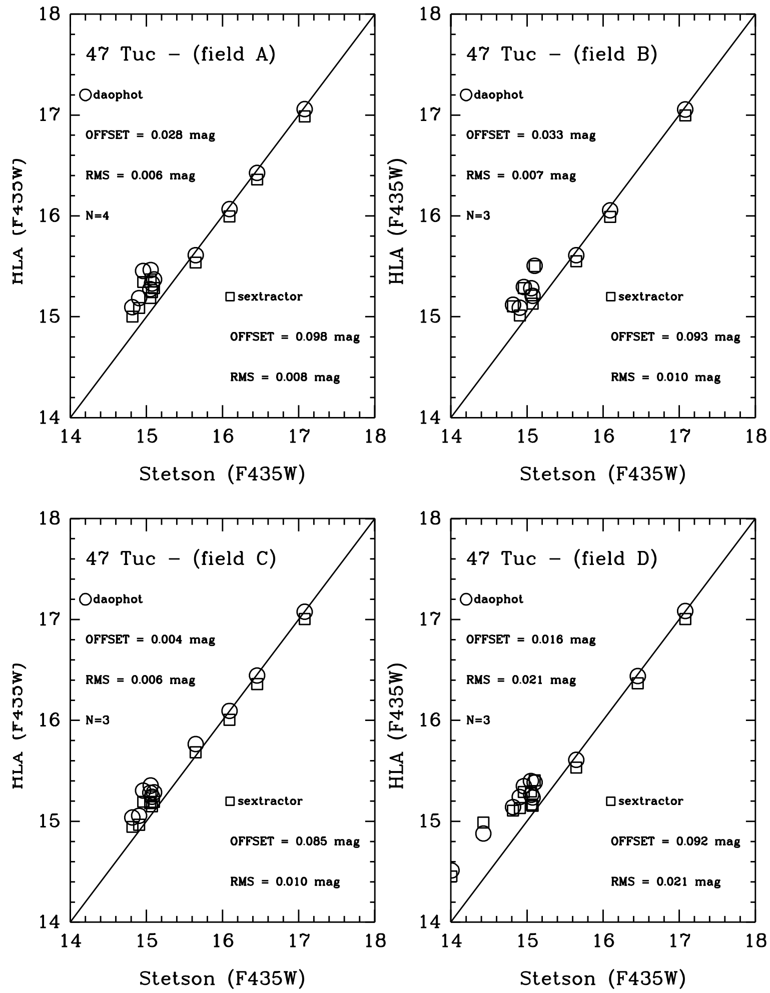

Figure # 3 – Comparison between HLA (DAOPHOT and Sextractor) photometry with Stetson ground-based photometry for 4 fields in 47 TUC. Typical offsets are a few hundredths of a magnitude with RMS < 0.01 mag. The stars on the left are saturated and hence not used in the comparison.

2. *Nearly point source photometry* (e.g., compact clusters in external galaxies) [Potential working group members: Brad Whitmore, Nate Bastian, Soeren Larsen, ...]

3. *Extended object photometry and morphology* (e.g., faint galaxies in survey fields, photometric redshifts) [Potential working group members: Harry Ferguson, Anton Koekemoer, Mark Dickinson, ...]

4. *Multi wavelength (and spectroscopic) cross matches* (e.g., ULIRGs, gamma-ray bursts, x-ray counterparts) [Potential working group members: Lee Armus, Bob Hanisch, Knox Long, ...]

5. *Time-resolved phenomena* (e.g., Cepheids, supernova, variable stars, …) [Potential working group members: Stefano Casertano, Adam Riess, Nathan Smith, …]

The working group members will serve primarily as consultants, and do most, if not all, of their work via e-mail and telecons. As the HSC is developed, it will be compared with existing catalogs, especially those of the working group members (e.g., see Figure 3). The comparisons will be published in the paper described in Section 8.

7. LEARNING FROM OTHER CATALOG PROJECTS

While each catalog has its unique set of challenges, a great deal can be learned by examining how other groups have addressed some of the issues raised in this white paper. The working groups discussed in section 6 will provide some of this knowledge, but it will also be necessary for the HSC project to proactively seek out this information for the most relevant examples (e.g., SDSS, Chandra, Spitzer, …)

This will be accomplished in two ways: 1) The project lead (and appropriate specialist within the group) will open a dialog with the key people from the other catalog projects, 2) various members of the HSC project will become active researchers using the various catalogs. This will include using the documentation and interfaces to see what works and what could be improved.

8. TOOLS

The utility of a catalog can be greatly enhanced if it is: 1) easy to access, 2) possible to "tune" for a user-specific purpose, 3) easy to cross-match with a user-provided catalog, and 4) possible to do quick-look analysis.

With this in mind, the HSC project will develop additional tools to extend and complement the existing HLA tools (e.g., footprints, interactive display, catalog overlays, simple line plotting, ...). These would include:

- Developing a relational database search capability (e.g., all Sc galaxies within a distance of 10 Mpc).

- Building user-defined, subset catalogs. The ability to select sources from the HSC according to uniform criteria, from anywhere on the sky, will help mitigate cosmic variance, which is one of the big issues of the traditional approach of doing contiguous surveys in just one part of the sky.

- Cross matching with existing user-provided catalogs.

- Plotting and quick-look analysis of time-resolved phenomena (e.g., search for possible periods).

- A three-plane visualization tool (i.e., where the three planes are image - plot - catalog: e.g., select "outliers" in one plane and instantly see where they are, and the statistics involved, in the other two planes).

- Ability to make new user-specified catalogs for certain datasets (e.g., use the software used to generate the HLA catalogs to make deeper measurements - to the 5 rather than 10 sigma level for example).

9. GETTING THE WORD OUT

Science cannot be enhanced if no one knows about the HSC. Besides advertising in the normal ways (i.e., MAST, HLA, HST websites, STScI newsletter, AAS meetings, ...) we will also have bi-annual workshops at STScI (tentatively the Fall of 2013 and 2015) titled "Doing Science with the Hubble Source Catalog". A PASP paper that provides the primary reference for the HSC project will be written. The paper will include a section where it demonstrates to what degree results from various research papers can be reproduced (e.g., to a completeness level which is 1.5 magnitudes shallower than a given study) from the HSC alone.

10. SUMMARY

An HST Source Catalog will add a valuable new dimension to the information available to both present and future users of the Hubble Space telescope for decades to come. It will enable efficient studies of time-variable phenomena and it will provide a bridge to future missions both in space (e.g., JWST), and on the ground (e.g., LSST).