Scientific Justification

Accretion is the fundamental process of cosmogony. In a universe dominated by cold dark matter in which structure grows gravitationally, large gravitationally bound objects form by accreting smaller ones. These bound objects—“dark matter halos”—contain galaxies, so we expect these accretion events to be followed by, or associated with, galaxy–galaxy interactions and then mergers.

The merger rate density for dark matter halos is known precisely (if CDM is the correct picture for structure formation), but precise predictions for galaxy–galaxy mergers have been hard to make in the context of a cosmological simulation, involving as they do baryonic and small-scale dark matter physics. Some approximations already exist and ambitious numericists expect in the near future to perform full-up computations with the required accuracy to (at least) predict rate densities.

In addition, theories predict precisely the distribution of halo–halo mergers across environment. Given our current constraints on how halos must be occupied by galaxies if CDM is correct (*e.g.,* Berlind et al., 2003), we can make strong predictions — once galaxy–galaxy merger rates can be predicted — for the environment-dependence of the galaxy–galaxy merger rate density. Indeed, there might be some strong “differential” predictions possible even now.

Oddly, the rate density of galaxy–galaxy merger events has been surprisingly hard to measure empirically. This difficulty exists in part because large, well defined samples are hard to assemble, in part because tidal features and other morphological signatures of merging are hard to identify automatically and consistently, and in part because we don’t in many cases know the time period over which a merger or merger remnant remains identifiable.

We can’t solve all of these problems in the next year or two! However, we can help by matching up our custom-made statistical sample of galaxies from the Sloan Digital Sky Survey (SDSS)—the 693,319 member NYU Value-Added Galaxy Catalog (NYU-VAGC)—with the HST archive, providing high-quality imaging for a large and (if we are careful enough) statistically unbiased sample of galaxies for which we have redshifts, high S/N spectra, and measures of environmental density. *Preliminary tests suggest that there will be HST imaging coverage for 2 to 3 percent of the NYU-VAGC sources, so our SDSS Data Release 2 data set will have up to 20,000 entries; our SDSS Data Release 4 data set will have up to 40,000.*

Our primary (but optimistic) goal is the absolute merger rate density. However, even when merger timescale uncertainties are large, with this large data set, the dependence of the merger rate on environment might be accurately measureable, because differential experiments can be constructed when we have enormous samples like the NYU-VAGC.

In the process, we will create an instantly public dataset (read “no proprietary period”) that provides redshifts, spectra, and environmental densities to HST imaging observations, and provides high resolution HST imaging to the massive SDSS data set.

With a small amount of continuing support, this dataset can be maintained and enlarged as more SDSS and HST data become available.

1 Post-starburst galaxies

By looking for excess A stars (relative to O and B stars) in galaxy spectra, we can identify galaxies for which it has been less than 1 Gyr since some terrifying event that caused a catastrophic drop in star formation. These "K+A" or "E+A" galaxies (Dressler & Gunn, 1983; Zabludoff et al., 1996; Goto et al., 2003; Quintero et al., 2004) are plausibly not only post-starburst but post-interaction, since it is expected that the event that shut off star formation involved a galaxy-galaxy interaction or merger. Indeed, if disk-dominated galaxies merge to form bulge-dominated galaxies, a central starburst is necessary to adjust the chemical abundances, remove the cold gas, and enhance the central stellar density. In particular, bulges are alpha-enhanced relative to disks (Worthey, 1998), lack star-formation fuel (Roberts & Haynes, 1994) and are higher in stellar density (as clearly seen in the SDSS, Hogg et al., 2002; Blanton et al., 2003a).

We have found a large class of post-starburst galaxies in the NYU-VAGC; they form a clearly separate population in spectrum space among the SDSS Main Sample galaxies (Quintero et al., 2004). To summarize: we were able to look at their mean environment (similar to disk-dominated galaxies, not as high density as bulge-dominated galaxies), and their surface brightnesses and radial profiles (similar to bulge-dominated galaxies, though their colors are bluer because of the A stars). All of this evidence strongly suggests that the post-starburst galaxies at $z \sim 0.1$ are the progenitors of field (*i.e.*, low density environment) ellipticals and S0s. Because the A stars have known lifetimes, we were able to use the A star abundance statistics among the post-starburst galaxies to compute a rate density for the events (mergers, perhaps?) that create them; we found a comoving rate density corresponding to about one percent of the L^* galaxy population per Gyr at $z \sim 0.1$ (for more details see Quintero et al. 2004).

There is some evidence that the roughly 10^3 (and now about twice that many with data that will be part of SDSS DR4) K+A galaxies found in the NYU-VAGC have unusual numbers of close neighbors. This evidence suggests that these galaxies may be involved in mergers. However, our ability to ask detailed questions about the events "causing" the K+A population is severely limited by the quality of the SDSS imaging, which only have kpc resolution imaging for typical targets.

Indeed, when Yang et al. (2004) observed a set of five very strong K+A galaxies with HST (their selection criteria being more stringent in A-star excess than ours), they found that all five showed some evidence of interactions, and that the "stronger" K+As (either younger or having a larger star formation event) had more obvious signs of interaction. This result alone motivates performing a full match of HST imaging with the NYU-VAGC, to extend the sample to less extreme post-starburst galaxies, to quantify trends, and to develop different kinds of control samples.

There are some uncertainties about the ages and timescales of some kinds of morphological indicators of interactions. Interestingly, if we can find tidal features in the post-starburst population, we might be able to attach to them a timescale using the time since burst inferred from the stellar populations. Unfortunately, with K+A spectra, there is a pretty strong degeneracy between the size of the burst and the time since the burst, although statistically there is a strong negative relationship between the abundance of A stars and the mean time since the burst, so there is hope that the trends of tidal features with stellar populations might "calibrate" some timescales. At the very least, tidal features that appear in K+A galaxies must have ages or timescales < 1 Gyr.

2 Galaxy pairs with tidal arms

Interacting galaxies have been prime targets for high-impact HST imaging. Think “The Antennae” (Whitmore et al., 1999), “Stephan’s Quintet” (Gallagher et al., 2001), and, soon, “M51”. However, it is a testament not just to the power of HST but also to the relative scarcity of clear-cut galaxy–galaxy mergers that we see these same few images displayed in seminars over and over again.

For some of these systems, the dynamics of the tidal arms can be reverse-engineered, and the time since the traumatic event can be robustly inferred. However, inferring a galaxy–galaxy merger rate from these observations is difficult, for a number of reasons, including

1. the lack of a well-defined “parent” sample from which they are chosen,
2. the lack of well-defined selection criteria, *e.g.*, criteria that could be implemented by an automated selection system,
3. incomplete understanding of the conditions that the galaxy orbits and galaxy dark-matter distributions must obey in order to raise large-scale tidal arms, and
4. the lack of (admittedly non-trivial) modeling to determine the time period over which the systems would be identified as ongoing mergers, by eye or by some automated system.

While we cannot solve all of these problems, we propose here to fix the first and ameliorate the second by associating all possible extant HST imaging with the enormous (and enormously voluminous) NYU-VAGC, and by post-processing the outputs of straightforward image fitting to—at the very least—the bulge-dominated galaxy population.

To simply assemble all the data and meta-data (see “Analysis Plan” section, below) into one place will make studies of the abundances and characteristics of merging and disturbed morphologies much more straightforward, with massive numbers of targets to investigate.

It is worth noting that the heterogeneity of the HST data, taken as they are through different filters for different exposure times, with different observing strategies, will make some kinds of analysis non-trivial. In particular, it is difficult to make automated measures of morphological asymmetries (*e.g.*, Odewahn et al., 1996; Conselice et al., 2000; Odewahn et al., 2002; Moustakas et al., 2004) and disturbances truly independent of obtained S/N ratio. However, it is worth keeping in mind that the normal galaxies in the NYU-VAGC are at redshifts $z \sim 0.1$, and enormous science has been produced with $ugriz$ imaging that represents 55 s exposures on a ground-based 2.5 m, so we could even *degrade* (heavens, no!) the best quality imaging data to the worst included S/N for the sample under consideration.

There are additional issues relating to the fact that different imaging programs work in different bandpasses. In the case of the bulge-dominated population, the galaxy color is close to constant from pixel to pixel. Therefore, any optical images in all bandpasses are linear scalings of the stellar surface density. This fact allows us to compare two such galaxies observed in different bandpasses in a fairly straightforward way. For more complex galaxies, such as spirals and star-forming galaxies, the colors vary dramatically across the galaxy, so this procedure is not so easy, and results must be given as a function of bandpass.

Finally (we hate to admit this in print), it ought to be noted that in fact it will be possible for us to manually inspect *all* of the HST observations in the NYU-VAGC-HST *by eye.* Though the NYU-VAGC contains 693,319 sources (and will contain many more when we upgrade to SDSS DR4), the total “footprint” of HST imaging on top of this sky area is expected to be only 2 to 3 percent, and we expect no more than 20,000 to get HST imaging. So although this hand-inspection task is large, and although this brings us back to classical, subjective morphological classification, it is a useful reminder that it will be possible to do great science with this data set even in the event that automated selection methods fail us.

3 Shells

Interestingly, although they are less dramatic in their appearance than tidal tails, shells in old, bulge-dominated galaxies may provide some of the cleaner and more reliable measures of interaction rate densities. There are a number of advantages, including that

1. shells probably trace the “dry merger” population, if indeed there are interactions that do not trigger any star formation, so they are somewhat orthogonal to post-starburst methodologies,
2. the parent galaxies tend to have simple morphologies and are transparent,
3. shells are relatively common, allowing the construction of samples with good statistics,
4. shells have a limited range of morphologies, unlike tidal arms, which can have a wide range of properties depending on age, viewing angle, and orbital parameters, and
5. because both the galaxies and the shells are made up of old stars (primarily), it is possible (in principle, at least) to estimate their lifetimes in models without the complexities of inhomogeneous star formation and fading.

In practice, we don’t yet know if these advantages will be sufficient to permit a fully automated, statistical study of shell galaxies. However, we have hopes for finding tangential edges in smooth-model–subtracted residuals in the images of early-type galaxies (see the “Analysis Plan” section, below).

One extremely tantalizing possibility is that the ratio of the shell radius $R_{\rm shell}$ to the velocity dispersion σ_v of the stars in the galaxy will provide a time $t_{\rm event}$ since the event that created the shell (Quinn, 1984). Of course we know from simulations that this is not going to be an accurate measure of time since the burst, and that multiple factors are involved in setting a particular shell’s radius and other properties (Hernquist & Quinn, 1988, 1989). However, there might be significant correlations between this time estimate and the true time, which can only be found in large statistical studies like those enabled by this Legacy AR program. In particular, comparison of shell existence and radii with the abundances of A stars (which evolves monotonically since the “event”) in post-starburst galaxies might—if we are extremely fortunate—produce an “evolutionary sequence” and constrain, calibrate, or rule out the use of shells as clocks.

Even without such good fortune, the creation of a large, easy-to-use data set relevant to shell galaxies is going to be very productive for studies of galaxy formation over the next decade.

4 HST imaging of the NYU-VAGC

The specific deliverable of this proposal is a marriage of all HST imaging to the entire NYU-VAGC to create the (badly named) NYU-VAGC-HST. This catalog—which has enormous numbers of legacy uses beyond the specific search for tidal features discussed in this proposal—will consist of the entire NYU-VAGC plus, where there is overlap, the HST imaging of NYU-VAGC galaxies and quasars. From preliminary tests, we predict that our combined catalog will have no more than 20,000 sources for SDSS DR2, and 40,000 for SDSS DR4 (coming soon).

The NYU-VAGC-HST provides

1. much higher resolution imaging for NYU-VAGC galaxies,
2. higher S/N imaging for NYU-VAGC galaxies,
3. new wavelength coverage for NYU-VAGC galaxies, in some cases,
4. uniform (though low resolution) $ugriz$ imaging for sources in a significant fraction of all HST imaging, and
5. redshifts for thousands of sources in HST imaging.

Our primary science goals benefit most from the first two.

The HST extension to the NYU-VAGC will greatly improve the scientific power in areas such as

1. shells, tidal arms, and other signs of interactions and mergers,
2. galaxy morphological mix as a function of environment, star formation rate, and other properties at all SDSS-accessible redshifts,
3. testing and validation of results based on the lower resolution and lower S/N SDSS imaging,
4. the geometry and extinction of dust in external galaxies,
5. the low-redshift structure along the line of sight to high-redshift HST targets for, *e.g.*, lensing effects, and
6. the existence and properties of quasar host galaxies (yes, any quasars selected for SDSS spectroscopy are included in the NYU-VAGC).

Our primary science goal is the first, and we intend to work on the first four, as they represent extensions of our ongoing galaxy properties projects (Hogg et al., 2002, 2003; Blanton et al., 2003a; Hogg et al., 2004, 2005b; Blanton et al., 2005a, 2004b).

We will assemble and make immediately public all of the data in the HST match to the NYU-VAGC, along with all the flags and associated information ("meta-data"; see the "Analysis Plan" section, below) that an investigator will need in order to perform statistical studies with the catalog so created.

References

Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, ApJ, 600, 681

Baldry, I. K. et al. 2005, MNRAS, in press (astro-ph/0501110)

Becker, R. H., White, R. L., & Helfand, D. J. 1995, ApJ, 450, 559

Berlind, A. A., Weinberg, D. H., Benson, A. J., Baugh, C. M., Cole, S., Davé, R., Frenk, C. S., Jenkins, A., Katz, N., & Lacey, C. G. 2003, ApJ, 593, 1

Blanton, M. R., Eisenstein, D. J., Hogg, D. W., Schlegel, D. J., & Brinkmann, J. 2005a, ApJ, in press (astro-ph/0310453)

Blanton, M. R. et al. 2003a, ApJ, 594, 186

Blanton, M. R. et al. 2003b, ApJ, 592, 819

Blanton, M. R. et al. 2004a, AJ, submitted (astro-ph/0410166)

Blanton, M. R. et al. 2004b, ApJ, submitted (astro-ph/0411037)

Blanton, M. R. et al. 2005b, ApJ, submitted (astro-ph/0410164)

Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., & Brinkmann, J. 2004, MNRAS, 351, 1151

Colless, M. et al. 2001, MNRAS, 328, 1039

Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886

de Vaucouleurs, G., de Vaucouleurs, A., Corwin, H. G., Buta, R. J., Paturel, G., & Fouque, P. 1991, Third Reference Catalogue of Bright Galaxies (Volume 1-3, XII, 2069 pp. 7 figs.. Springer-Verlag Berlin Heidelberg New York)

Dressler, A. & Gunn, J. E. 1983, ApJ, 270, 7

Eisenstein, D. J. et al. 2005, ApJ, submitted (astro-ph/0501171)

Gallagher, S. C., Charlton, J. C., Hunsberger, S. D., Zaritsky, D., & Whitmore, B. C. 2001, AJ, 122, 163

Glazebrook, K. et al. 2003, ApJ, 587, 55

Goto, T. et al. 2003, PASJ, submitted (astro-ph/0301305)

Hernquist, L. & Quinn, P. J. 1988, ApJ, 331, 682

Hernquist, L. & Quinn, P. J. 1989, ApJ, 342, 1

Hogg, D. W., Eisenstein, D. J., Blanton, M. R., Bahcall, N., Brinkmann, J., Gunn, J. E., & Schneider, D. P. 2005a, ApJ, in press (astro-ph/0411197)

Hogg, D. W., Tremonti, C. A., Blanton, M. R., Finkbeiner, D. P., Padmanabhan, N., Quintero, A. D., Schlegel, D. J., & Wherry, N. 2005b, ApJ, submitted (astro-ph/0408420)

Hogg, D. W. et al. 2002, AJ, 124, 646

Hogg, D. W. et al. 2003, ApJ, 585, L5

Hogg, D. W. et al. 2004, ApJ, 601, L29

Hoyle, F., Rojas, R. R., Vogeley, M. S., & Brinkmann, J. 2003, ApJ, submitted (astro-ph/0309728)

Hoyle, F., Vogeley, M. S., Gott, J. R. I., Blanton, M., Tegmark, M., Weinberg, D. H., Bahcall, N., Brinkmann, J., & York, D. 2002, ApJ, 580, 663

Moustakas, L. A. et al. 2004, ApJ, 600, L131

Odewahn, S. C., Cohen, S. H., Windhorst, R. A., & Philip, N. S. 2002, ApJ, 568, 539

Odewahn, S. C., Windhorst, R. A., Driver, S. P., & Keel, W. C. 1996, ApJ, 472, L13

Padmanabhan, N., Seljak, U., Strauss, M. A., Blanton, M. R., Kauffmann, G., Schlegel, D. J., Tremonti, C., Bahcall, N. A., Bernardi, M., Brinkmann, J., Fukugita, M., & Ivezić, Ž. 2004, New Astronomy, 9, 329

Pope, A. C. et al. 2004, ApJ, 607, 655

Quinn, P. J. 1984, ApJ, 279, 596

Quintero, A. D. et al. 2004, ApJ, 602, 190

Roberts, M. S. & Haynes, M. P. 1994, ARA&A, 32, 115

Rojas, R. R., Vogeley, M. S., Hoyle, F., & Brinkmann, J. 2003, ApJ, 617, 50

Saunders, W. et al. 2000, MNRAS, 317, 55

Shen, S., Mo, H. J., White, S. D. M., Blanton, M. R., Kauffmann, G., Voges, W., Brinkmann, J., & Csabai, I. 2003, MNRAS, 343, 978
Tegmark, M. et al. 2004a, Phys. Rev. D, 69, 103501
Tegmark, M. et al. 2004b, ApJ, 606, 702
Whitmore, B. C., Zhang, Q., Leitherer, C., Fall, S. M., Schweizer, F., & Miller, B. W. 1999, AJ, 118, 1551
Worthey, G. 1998, PASP, 110, 888
Yang, Y., Zabludoff, A. I., Zaritsky, D., Lauer, T. R., & Mihos, J. C. 2004, ApJ, 607, 258
Zabludoff, A. I., Zaritsky, D., Lin, H., Tucker, D., Hashimoto, Y., Shectman, S. A., Oemler, A., & Kirshner, R. P. 1996, ApJ, 466, 104
Zehavi, I. et al. 2002, ApJ, 571, 172

Analysis Plan

Because our specific science goals can only be achieved once our legacy data products are in place, we describe the data products before we describe our particular science plan.

1 The NYU Value-Added Galaxy Catalog (NYU-VAGC)

The currently public version of the NYU-VAGC (based on SDSS Data Release 2) is described in full by Blanton et al. (2004a); here we give a brief description of it, prior to discussion of the HST extension. The next incremental version (based on upcoming SDSS Data Release 4) will be ready soon; its contents will be similar, but for a larger total sample. Everything is strictly versioned and old versions are maintained in parallel.

The NYU-VAGC contains matches among the SDSS spectroscopic survey, 2MASS, the PSCz survey (Saunders et al. 2000), the Faint Images of the Radio Sky at Twenty-centimeters (FIRST; Becker et al. 1995) survey, the Third Reference Catalog of Galaxies (RC3; de Vaucouleurs et al. 1991), the Two-degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001), the Galaxy Evolution Explorer (GALEX), and the Spitzer SWIRE data.

What makes this catalog more than just a matched list of objects from all of these surveys? We have performed (or obtained from others) a number of unique sorts of measurements. In addition, we have performed a large number of quality control checks, and validated most of the cataloged quantities by publishing scientific analyses based on them. For example, for each object we provide

1. its local environmental density (by several different measures),
2. the appropriate K-corrections in the SDSS $ugriz$ bands and in the 2MASS JHK_s bands,
3. a Sérsic fit in each SDSS band,
4. the results of the SDSS spectroscopic analysis of Brinchmann et al. (2004),
5. GALEX fluxes in the UV, where available, and Spitzer/SWIRE fluxes in the near- and mid-IR, where available.

If the object happens to be nearby (*i.e.*, at very low redshift) we also

1. provide a correction to its redshift for peculiar velocities; and
2. check if it is a double star (many of which are incorrectly labeled as resolved galaxies by

the SDSS pipelines).

Often large galaxies are incorrectly interpreted by the SDSS software as a number of much smaller galaxies. In addition, this "deblending" sometimes makes different decisions for the latest data reprocessing than for the original, targeting version. In order to alleviate the bookkeeping problems this feature causes, we

1. use an improved algorithm to match spectra to photometric objects which is more sophisticated and more accurate than a simple positional match; and
2. check bright or nearby galaxies by eye to determine whether they have been deblended, and flag them as such in a quality flag.

In addition, we have created well-defined samples of galaxies useful for statistical and large-scale structure analysis. In particular, we have created a large-scale structure sample based on the SDSS with a full description of the window function, the completeness and flux limit in each direction, and random catalogs. The descriptions we distribute are the *only* such descriptions for SDSS data.

All of these improvements have been motivated by specific scientific goals, have been tested in the context of those goals, and (as described below) have resulted in scientific papers.

Topics under investigation with the NYU-VAGC in its current state include:

1. the correlation function of galaxies (Zehavi et al., 2002; Eisenstein et al., 2005) and power spectrum of galaxies (Tegmark et al., 2004b,a; Pope et al., 2004),
2. cosmic homogeneity (Hogg et al., 2005a),
3. topology of the galaxy distribution (Hoyle et al., 2002),
4. bispectrum of galaxies (Scoccimarro *et al.*, in preparation),
5. galaxy groups (Berlind *et al.*, in preparation),
6. void galaxies (Rojas et al., 2003; Hoyle et al., 2003),
7. properties of star-forming galaxies (Brinchmann et al., 2004), and post-starburst galaxies (Quintero et al., 2004),
8. properties of extremely low luminosity galaxies (Blanton et al., 2005b),
9. dynamics of elliptical galaxies (Padmanabhan et al., 2004),
10. cosmic spectrum of optical galaxy light (Glazebrook et al., 2003),
11. the bimodel color–magnitude distribution of galaxies (Baldry et al., 2004),
12. mid- and near-infrared properties of galaxies (Hogg et al., 2005b), and
13. the optical luminosity function and its evolution (Blanton et al., 2003b; Baldry et al., 2005),
14. the distribution of optical galaxy properties (Shen et al., 2003; Blanton et al., 2003a) and their relationships with environment (Hogg et al., 2003, 2004; Blanton et al., 2005a, 2004b)

Even though many of the above authors are among the 200 SDSS participants, they are spread over more than 30 institutions. *Our communication with users of the sample is primarily through data files and a thorough system of documentation.* Many of the users of the current NYU-VAGC are non-experts in SDSS technical information, and non-experts in SQL, databases, and data reduction. They are statistical astrophysicists who want to ask reliable statistical questions of the data *without* knowing everything about everything in the survey. The success of these users in doing good science with our data set is a testament to its good construction, scientific validity, and user friendliness.

1.1 The HST extension (NYU-VAGC-HST)

We expect to extract from the HST observations and keep

1. a list of the overlapping HST imaging for every NYU-VAGC galaxy, along with instrument, filter, date, and exposure time information, even when there is no detectable source in the HST imaging (we expect this to be rare),
2. "postage stamp" images cut out from the HST imaging, using the SDSS catalog and SDSS-measured sizes to set the cutout parameters (which we will make significantly larger than the galaxy itself, especially since tidal features are often at large radius),
3. fluxes measured in the HST imaging through "matched apertures" matched to the SDSS photometric methods, including the resolution differences between SDSS and HST, to make reliable Petrosian and "model" colors in analogy to the SDSS colors, with which we, and many others, have familiarity, and
4. the RA, Dec, redshift, and SDSS spectrum identification information for every SDSS spectrum that overlaps any HST pointing, for use by the HST community (and possible re-inclusion into the HST Data Archive).

Will the HST imaging of the NYU-VAGC represent a completely random subsample of the NYU-VAGC? Of course not. For example, if a certain sub-class of nearby galaxies was a target of a successful HST program, then that sub-class is more likely than an average galaxy of the same brightness and redshift to have been observed by HST. Furthermore, galaxies are highly clustered, so galaxies in high density regions are more likely than those in low density regions to be in the field of view of an HST pointing targeted at another galaxy at a similar redshift. Furthermore, HST pointings are not equally weighted towards all environments (think, *e.g.*, of all the cluster science programs). This makes it non-trivial to do correct statistical science with the NYU-VAGC-HST.

Nonetheless, it is possible to use data with complicated selection criteria in statistical studies, provided that sufficient information is kept. Indeed, our primary interest is in statistical studies. In addition to the extracted data listed above, we will also keep "meta-data" including

1. the name, RA, Dec, redshift (where known), and type (*e.g.*, star, galaxy, quasar, *etc.*, where known) of the primary target of the HST imaging from which the imaging data for each NYU-VAGC galaxy was extracted,
2. a flag indicating whether or not, to the best of our judgement ("don't know" will be an option) the NYU-VAGC galaxy in question was the explicit target of the HST imaging, and
3. the redshift difference, to the best of our judgement ("don't know" will be an option), between the NYU-VAGC galaxy in question and the target of the HST imaging.

When we are doing very sensitive statistical tests, we will restrict our use of the NYU-VAGC-HST to only those HST data for which the meta-data indicates that the NYU-VAGC galaxy was not the target of the imaging, and also not even at a nearby redshift. For other studies, where spanning representative environments can be shown to be unimportant for the purpose (*e.g.*, for a study of galaxies already pre-selected to be in certain types of rich clusters), we need only restrict to HST data for which the NYU-VAGC galaxy was not the target. For other tests, even that is not necessary.

All of these meta-data will be available to all investigators; indeed it will be necessary for their statistical work.

Why can't we just submit the 693,319 RAs and Decs to the MAST interface and be done with it? There are several important—and time-consuming—respects in which this project goes way beyond a MAST query.

1. On a relatively trivial note, we can't do any MAST query this large in one shot, so we have to automate (in a non-destructive way), a large number of independent MAST queries.
2. MAST does simple cone searching; it does not do detailed instrument footprints, so we have to post-process MAST outputs to determine whether an object actually was within the field-of-view of a given observation.
3. We have to extract the HST imaging data from the calibrated data in MAST to make the postage stamps.
4. We perform SDSS-like and resolution- and aperture-matched photometry on the cutouts.
5. MAST cannot return the necessary meta-data (described above) we need for statistical studies with the outputs.

Indeed, these limitations of MAST are probably under discussion at MAST, and enhancements like these may be possible for the future. *We consider this project as a first step in developing new tools that will be generally useful, to MAST, and to future VO-like projects.* All software we write is open-source and freely distributed by us.

1.2 Data release

As with all the work we do on the NYU-VAGC, all data we compile, and all trustworthy measurements of the data we make, will be made fully public immediately in the NYU-VAGC, with no proprietary period. In addition, we will be happy to mirror the NYU-VAGC in the MAST archive at STScI, if that is of interest to MAST and the community.

2 The galaxy–galaxy merger rate density

We already perform by-eye inspections of large numbers (tens of thousands) of SDSS galaxies, to do morphological classification and look at signs of interaction. We are therefore prepared, tool-wise and mentally (!) for visual inspection of all outputs.

We will compare incidence of tidal/disturbance features of different kinds with the young star fraction in post-starburst galaxies, to see if we can discern a time-evolution of the features that coincides with the time evolution of the young star fraction.

We already perform automatic fitting of ellipsoidal Sérsic models to the SDSS imaging data. We will adapt the (open-source, public) code we have developed to automatically fit all HST images of NYU-VAGC galaxies. The residuals to these fits will show

1. spiral structures,
2. HII and star-formation regions,
3. tidal features, and
4. shells.

We will implement automated asymmetry measurement on these residuals to flag galaxies for further inspection; we will inspect these by eye.

Finally, shell features tend to show tangential "hard" edges in the residuals. We will implement simple local radial/tangential transforms that look for hard tangential edges in an angular sector of a galaxy profile. We will use well-known shell galaxies (such as NGC 5128, NGC 474, *etc.*), artifically "moved" to higher redshift, to tune the parameters of our automated search.

We will use SDSS-measured velocity dispersions to compute, along with shell radii, rough timescales, and look at whether the timescales so found correlate in reasonable ways with other possible post-merger indicators (such as stellar populations or tidal tails, etc). If all goes well (many "if"s), we will work towards computation of a dry merger rate density in the $z \sim 0.1$ Universe.

Budget Narrative

We propose to do all work on this proposal in a single, high-intensity year.

One summer month is provided for Hogg, who will perform the comparative data analysis of the post-starburst galaxies and the control samples, and who will supervise the junior postdoc.

Six months are provided for Blanton, who will make final decisions about what data and what flags and supplemental information needs to be included in the catalog, who will implement the automated shell detection, and who will supervise the undergraduate researcher.

One year is provided for a junior postdoc, who will work on converting the shell information and other results into statistics relevant to the galaxy–galaxy merger rate density, and analyze the uncertainties, and, if possible, relevant simulation or theoretical outputs.

One year (full-time summer and part-time year) is provided for an undergraduate researcher, who will manage the HST data identification and download, meta-data collection, postage-stamp making and photometry, under the supervision of Blanton.

A fast, high-RAM, large-storage server machine is provided to ensure that data flow and staging does not become the limiting factor for the personnel.

Travel for all the personnel to two domestic meetings is also provided, along with publication and dissemination costs (*e.g.*, presentation software, poster printing, and page charges) to support the dissemination of the results.

Previous Related HST Programs

While both of the investigators have several successful HST programs, none are directly relevant to the current proposal.