

Creating updated, scientifically-calibrated mosaics for the RC3 Catalogue

1. Introduction

Astronomical catalogs such as the famous Messier Catalog and New General Catalog (NGC) has its historical usage in helping astronomers who are studying objects that shares some common property. This was particularly important in the days when all-sky information was not available, so data was taken by pointing the telescope at some particular location. Wide sky surveys such as the Sloan Digital Sky Survey (SDSS), Digitized Sky Survey (DSS), and Two Micron Sky Survey (2MASS) not only provide improved astrometry for finding these objects, but also photometric (other info?) on detected sources, such as those generated through SDSS's photo pipeline (CITE). With large sky surveys such as the SDSS, the job of target-based imaging is a lot easier since the position values of objects of interest are readily accessible via NED/ SIMBAD. The role of the astronomical catalogue has evolved into sample for statistical studies for some particular type of object (QSO, pulsars, LRG...etc) or cosmology studies.

The Third Reference Catalog of Bright Galaxies (RC3) compiled by de Vaucouleurs et al. [7], contains a complete listing of 23,022 galaxies with D_{25} apparent major isophotal diameter greater than 1 arcminute and with a total B-band magnitude greater than 15.5 nanomaggies.

The RC3 Catalog is an update of the original Reference Catalog of Bright Galaxies. It is `TALK ABOUT DATA`, `ViZier`. `HOW OFTEN UPDATED..etc?`

There has been existing literature on making mosaics for RC3 galaxies using data from the Sloan Digital Sky Survey (SDSS). Hogg and Blanton made g,r,i colored images of selected RC3 galaxies by using data from the SDSS DR6.¹ The EFIGI catalog (Baillard et al. [2]), dedicated to studying morphology of galaxies, also obtained a subset of 4458 RC3 FITS and color images using SDSS DR4 data and discarded artifacted or missing data via visual inspection. We perform the mosaicing procedures on the whole RC3 catalog using an automated positional update algorithm on the most recent SDSS DR10 imaging data. In addition, we designed a mosaicing pipeline that creates scientifically-calibrated mosaics, color-band images, as well as a set of updated positions, which can be easily adapted for existing surveys such as Two Micron All Sky Survey (2MASS), The Palomar Digital Sky Survey (DPOSS), or future imaging data release from Dark Energy Survey

(DES) and Large Synoptic Survey Telescope (LSST) for larger sky coverage and multiband images.

2. Data

3. The RC3 Catalog

Since the RC3 catalog is a reasonably complete representation of large, bright nearby galaxies in the extragalactic sky, its value is still evident in its usage in literature. Selected galaxies or complete subsets are used in astrophysical studies of quasars and X-ray sources or galaxy morphology and properties (CITE). The RC3 catalog is also statistical studies for cosmology purposes. Cross correlated also used in Updated Catalog, such as the New York University-Value Added Galaxy Catalog (NYU-VAGC)

3.1. Positional Inaccuracy in RC3 Catalog

Original Reference Catalogue of Bright Galaxies (CITE) uniform update "collected from 150 different sources", each telescope has different accuracy precision..etc The principle motivation behind the automation algorithm is to resolve the problem of centering galaxies and finding RC3 galaxy in a given field of view. Initially, using the standard mosaicking steps in Montage, we obtained many images with off-centered or missing galaxies due to the inherent inaccuracy of the positions in the catalog. There is a non-uniform update of the catalog over the years (—) after the catalog was published, more accurate position was collected from whatever sources/data they had access to. The B2000 coordinates (FK4 frame) in the RC3 catalog are denoted with two different levels of accuracy: HH MM SS.s, DD MM SS for the positions that has been updated (with accuracy of about 5-8 arcsec) and HH MM.m, DD MM for galaxies whose positional accuracy remain 1-2 arcminutes in the original catalog. In the latest version of the catalog (de Vaucouleurs et al. [6]), there remains 5492 RC3 galaxies that fall in the latter group.

3.2. Data

3.2.1. SDSS

SDSS is a optical 2.5m telescope at the Apache Point Observatory in New Mexico. We use the imaging data from Data release 10 which is part of the SDSS-III. SDSS-III is dedicated to studying galaxy clustering by looking at signatures of Baryon Acoustic Oscillation, structure and history of the galaxy, and the population of exoplanets.

¹<http://cosmo.nyu.edu/hogg/rc3/>

Feeding the SDSS data into the mosaicing pipeline has several advantages over data from other surveys, 1) It is one of the newest all-sky survey so has improved instrumentation and telescope compared to DPOSS. 2) On a related note, the resolution of SDSS is better than 2MASS and DPOSS. 2MASS has a “Tile” size, each Tile, so each pixel represents —degree, which is much larger than SDSS’s pixel size of —. SDSS imaging is also superior to DPOSS since its CCD-based imaging does not suffer from the same boundary edge distortion (WHAT IS THE NAME FOR THIS??) typical artifact of photographic plate imaging. 3) SDSS is an optical survey, so obviously “looks” much better than 2MASS’s infrared data since mapping g,r,i band to R,G,B is approximately accurate in value. However when mapping 2MASS’s J,H,K band, the relative values show up (which is longest wavelength and shortest), so the contrast in feature is shown. But it is not true color.

(TALK ABOUT HOW SDSS IMAGING DATA IS PROCESSED; SEE PHOTOLITE PAPER LUTPON) The SDSS data is pass through —pipeline, deblending, source, SKy-Server provide many info

3.3. 2MASS

The Two Micron All Sky Survey (2MASS; CITE) is a infrared survey in the J(1.25 mm), H(1.67 mm), and K(2.17 mm) bands. 1.3m telescope in (Does 2MASS have more advantage in seeing “cooler” object that give off thermal radiation?) - problem with low resolution large tile background - The benefit of smoother larger pixels (why?)

3.3.1. POSS-II

The Second Palomar Observatory Sky Survey (POSS-II) was a photographic survey that covered most of the northern night sky in the B, R, IR bands. It is a second generation survey 1980s update of POSS-I which was one of the first major all-sky photographic survey. We also serve FITS mosaics from POSS-I which contains only the B and R band data taken in the 1950s to serve as comparison. This data can also be joined with another single band survey and create colored mosaics..

The photographic plates used to create our data product are made accessible through STScI’s effort in Digital Sky Survey (DSS) project. large photographic plates

4. Automation Program

4.1. Algorithm

To address the problem of inaccurate positions discussed in section (?), we develop an algorithm that first updates the position by “finding” the galaxy in the imaging data. An RC3 object contains its The Catalogue of Principal Galaxies (PGC) number, updated and catalog ra,dec,radius for each galaxy in the RC3 catalog. Since every RC3 galaxy has a unique PGC number, we record its PGC number in order to identify the RC3 galaxy of interest for cases of source confusion. The method `source_info`

updates the position of individual RC3 objects by first creating a single band mosaic with a field of view 6 times the radius of the galaxy, then SExtractor detects all sources in the mosaic and select only the ones with a radius greater than 5.94 arcseconds. If the mosaicking field is chosen correctly, then SExtractor’s skylevel estimation is fairly accurate. Nevertheless, this size cut is large enough such that it eliminates most of stellar point sources and mistaken background noises but retains the subset of RC2 galaxies inside the RC3 Catalog that are smaller than 1 arcminute as described in de Vaucouleurs et al. [5].

If there are multiple galaxies in the field of view that satisfy this criteria, then it is passed into the source confusion algorithm to verify which one is the galaxy of interest. Then we record the single source of interest and generate mosaic FITS file for all bands and color images. If no sources are detected, then we mosaic the single band FITS using a larger field of view. This is done recursively, while keeping track of the number of iterations in each recursive steps. The process is terminated at the 5th recursive step.

4.2. Source Confusion

(One line intro about galaxies living close to each other== galaxy cluster, gravity stuff?) The source confusion algorithm works on the assumption that the only galaxies that are large enough to cause source confusion also lie in the RC3 catalog, since the catalog is reasonably complete for galaxies having apparent diameters larger than 1 arcmin. First, the program retrieves a list of n number of RC3 galaxies that may lie in the field of view of the generated mosaic. The `otherRC3` method in the `Server` class queries this information using the Vizier Catalog. If this information is provided by a survey’s query service (such as the RC3 Table in SDSS’s SkyServer), it can be overridden. The goal is to match the results returned by the server with the list of n largest sources detected by SExtractor.

We need to cross-correlate the coordinate by matching together the relative differences between the sources, due to the unreliability of absolute coordinates inherent to the catalog as discussed in section 2.1. We assume that the positional inaccuracy is due to instrumentation and measurement error but retains the galaxies’ relative locations. Therefore, we compute all the possible distances between any two RC3 galaxies that lies in our field. Then compare this with the set of differences generated by the SExtractor-generated list. We select n number of galaxies that has the closest values in both list, thereby matching together the galaxies from the catalog with the detected sources. Finally, we mosaic using that galaxy’s coordinates as center. Our premises proves to be valid, as the algorithm executed with a success rate of 99.97% in the SDSS run, and is capable of correctly resolving up to 8 RC3 sources in one example.

5. Pipeline

5.1. Class Hierarchy

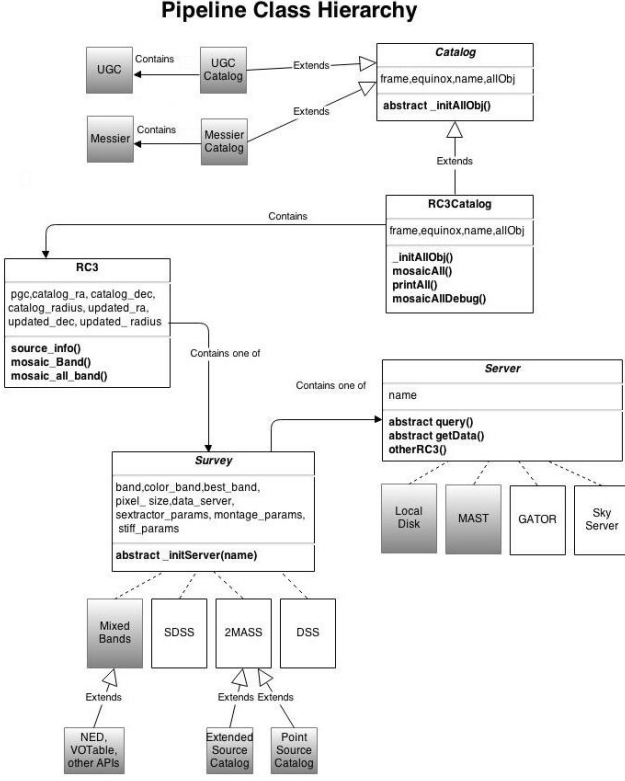


Figure 1: Greyed out boxes are possible extensions of this pipeline

The motivation behind building a class hierarchy is to establish a mosaicking pipeline for any given catalog and survey data. We identify the essential components required for mosaicking program to work. The relationship between these essential steps are reflected in our design of the abstract `Survey`, `Catalog`, and `Server` classes.

Server. Data acquisition is composed of two main tasks: querying imaging data and retrieving data from server. Most surveys have an API that enables data access by SQL or customized query. To make the `Server` class work, we need to figure out how to convert positional values (ra,dec) to record-keeping parameters dependent on the survey’s telescope. (e.g. tiles, frames...etc) For example, SDSS image frames are uniquely identified by a particular combination of run, camcol, and field. This step can usually be done using the SQL search. Our choice of using a server instead of a data object enables code reusability across various surveys that uses common server tools, such as IPAC’s GATOR query service (Alexov et al. [1]) or Astroquery.

Catalog Objects. For the purpose of mosaicking RC3 galaxies, we did not create a `CatalogObject` class, although it would be helpful to have a generic class similar to the

functions in `RC3Objects`. They perform the essential mosaicking features on a per-object basis so that they can be further used in the `Catalog` class. The final step in the mosaic procedure generates 2 TIFF color images. One emphasizes on low surface structures so that the halos around the galaxy can be seen; the other image is a poster publication image that uses higher cuts on the background to ensure a “clean” contrasting image. When extending the pipeline to other surveys, investigators can adjust the STIFF parameters to better suit the telescope-specific imaging details following the guidelines in Bertin [3].

Catalog. The `Catalog` class contains a list of objects that lies inside the catalog. Although the use of the `Catalog` container class seem unnecessary, it enables a clear separation between basic mosaicking functionality for individual galaxies. This can be used to study some particular object, executed on all objects in a `Catalog`, or simply used for debugging purposes. This abstraction barrier ensures minimal changes to the code in both classes when an investigator decides to input data from a new survey in the future.

5.2. Technical Details

Due to the recent surge in popularity of Python in astronomy, we wrote the program in Python 2.7.6 so that the program’s dependencies are more widely supported for extending its usage on future datasets. Most of the mosaic steps are done using IPAC’s Montage [8] mosaic engine along with the AstroPy Montage wrapper² and the final colored image is created using Astromatic’s STIFF v.2.4 [3]. Our choice of the two program makes use of the best features from both programs. Montage creates scientifically-calibrated images by retaining the astrometry and photometry of input sources during image reprojection step. STIFF provides the flexibility of adjusting many variables for the final colored image, as well as automatically estimating upper and lower cuts on the dynamic range using statistics derived from a pixel histogram. The source extraction is done using SExtractor v.2.19.5 (Bertin and Arnouts [4]). The resulting database was created using the Python sqlite3 module and the web search interface was written with PHP.

5.3. Known Errors

(Need to fix this section later ..) Even though a series of exception handling and error prevention mechanisms were put in place, there are still errors in the data product that we produced.

- mProjectExec is the montage Mosaic procedure that creates the reprojected image from the raw FITS files. Sometimes reprojected images are not created even when Montages’ debug statement clearly shows

²<http://www.astropy.org/montage-wrapper/>

that the reprojection was successful and table and header files are corrupted. This results in an error in later mosaicing steps. We have implemented error prevention mechanism to ensure that mosaic procedures terminates correctly in such cases and wrote the problematic galaxy into `failed_projection`, which can be examined later.

- Montage’s background rectification module is not used in creating the FITS mosaics. Attempts were made in implementing these procedure, however it is sometimes stuck in the step where the background model is applied to the projected images and produces huge difference images files (50GB) for reason that we have not figured out yet. This should not affect the astrometric quality of the data product. The only visible difference is observed in the color images for 2MASS, so will likely affect the photometric quality of the resulting 2MASS mosaics.(?)
- (We should probably fix this?) There was more galaxies affected by the source confusion issue since we are only taking one photographic plate to crop, it doesn’t actually pass through the mosaicing algorithm. Subsequently, it was never passed through the source confusion algorithm.

5.4. Performance

We accelerate the mosaicing process by performing the recursive algorithm on only a single band FITS file designated as `best_band`, then mosaicing all bands only once per object. For SDSS, we use the image obtained by the r band filter since transmission curve in Stoughton et al. [11] shows that r band has the highest quantum efficiency.

Most of the computation time is spent on downloading the raw FITS files from the survey’s specific server. Therefore, even though Montage’s modular design enables its performance to scale with number of processor (Jacob et al. [8]), the process would not be significantly sped up by the use of Message Passing Interface (MPI). Therefore, the speed depends more on the bandwidth (downloading rate) rather than the number of cores in the machine. This average runtime would also depend on the sky coverage of the particular survey and the speed of query.

This process can be significantly sped up if the investigator already has imaging data stored locally on a disk or have access to running the mosaic program alongside the survey’s datacenter. We have designed our class hierarchy such that this can be written as a subclass of `Server` with user-defined details of where the images are stored.

5.5. Result from SDSS/DPOSS run

There are a total of 12512 RC3 galaxies within the SDSS footprint. The automated mosaicing pipeline was successful for 90.22% of the galaxies. 2446 RC3 galaxies is mosaiced using an updated position that was more than 1 arcminute off the values recorded in the RC3 catalog. On

(bigdog specs) machine, the program processes on average about 80 RC3 objects per hour using SDSS data. The finished data product occupies 31 GB of disk space.

For the DPOSS Run, on average 0.31 galaxies per minute. (#More on this later..)

6. Conclusion

Wide-area sky surveys are often used to answer fundamental questions regarding large-scale structure and the cosmological history of the universe. The mosaicing procedures described in this paper provides a convenient way to generate mosaics for sky survey imaging data for a given catalog, or on a per-image basis. This is easily adapted for future data to create the scientifically-calibrated FITS mosaics as well as colored images.

Feeding in the new data enables a greater sky coverage and higher resolution images of objects inside the catalog. The scientifically calibrated FITS images can be used as for individual studies, as well as used for improved astrometry in updating the catalog. It can also be combined with the calibrated mosaics of different wavelengths into multi-band color images with mosaics from single/double band surveys such as POSS-I and GALEX.

The data product may also be useful during the target selection/commissioning stage of new sky survey programmes such as the LSST. With the trend of future telescope having larger aperture(focal plane?), large nearby galaxies needs to be properly masked to prevent the loss of imaging details from saturated CCDs. The updated RC3 coordinates tells us which regions of the sky may be affected by these large galaxies, and the FITS files can be used for model-fitting the galaxy’s light distribution. The FITS images can be used along with existing tools for scientific analysis, such as Astrometry.net(Lang et al. [9]), SExtractor, APLpy(Robitaille and Bressert [10]). The program is designed such that other API with Programs interface such as HEASARC’s SkyView, NED ..etc can easily be latched on (REFERENCE Fig1). This can also be extended for other astronomical catalog such as the Messier Catalog or New General Catalog (NGC), as well as catalogs with user-defined selection limit.

We have provided documentation on GitHub that guides investigators through the steps to adapting the pipeline for future imaging data. The source code and documentation for the pipeline described in this paper can be found in the project repository: <https://github.com/ProfessorBrunner/rc3-sdss>. The data product from the SDSS and DPOSS run is available via a search form at <http://lcdm.astro.illinois.edu/search.html>.

Acknowledgements

(Google GSoC?) We thank Harold G. Corwin, Jr. for helpful discussion that helped this work. This research made use of Montage, funded by the National Aeronautics and Space Administration’s Earth Science Technology Office, Computation Technologies

Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology. Montage is maintained by the NASA/IPAC Infrared Science Archive. This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration, 2013).

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University

- [1] A. Alexov, G. B. Berriman, N.-M. Chiu, J. C. Good, T. H. Jarrett, M. Kong, A. C. Laity, S. M. Monkewitz, N. D. Tahir-Kheli, S. W. Norton, and A. Zhang. The NASA/IPAC Infrared Science Archive (IRSA): The Demo. In P. Shopbell, M. Britton, and R. Ebert, editors, *Astronomical Data Analysis Software and Systems XIV*, volume 347 of *Astronomical Society of the Pacific Conference Series*, page 7, Dec. 2005.
- [2] A. Baillard, E. Bertin, V. de Lapparent, P. Fouqué, S. Arnouts, Y. Mellier, R. Pelló, J.-F. Leborgne, P. Prugniel, D. Makarov, L. Makarova, H. J. McCracken, A. Bijaoui, and L. Tasca. The EFIGI catalogue of 4458 nearby galaxies with detailed morphology. 532:A74, Aug. 2011. doi: 10.1051/0004-6361/201016423.
- [3] E. Bertin. Displaying Digital Deep Sky Images. In P. Ballester, D. Egret, and N. P. F. Lorente, editors, *Astronomical Data Analysis Software and Systems XXI*, volume 461 of *Astronomical Society of the Pacific Conference Series*, page 263, Sept. 2012.
- [4] E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement*, 117:393–404, June 1996.
- [5] G. de Vaucouleurs, A. de Vaucouleurs, and J. R. Corwin. Second reference catalogue of bright galaxies. In *Second reference catalogue of bright galaxies, 1976*, Austin: University of Texas Press., page 0, 1976.
- [6] G. de Vaucouleurs, A. de Vaucouleurs, H. G. Corwin, Jr., R. J. Buta, G. Paturel, and P. Fouqué. *Third Reference Catalogue of Bright Galaxies. Volume I: Explanations and references. Volume II: Data for galaxies between 0^h and 12^h. Volume III: Data for galaxies between 12^h and 24^h*. 1991.
- [7] G. de Vaucouleurs, A. de Vaucouleurs, H. G. Corwin, R. J. Buta, G. Paturel, and P. Fouque. Third Reference Cat. of Bright Galaxies (RC3) (de Vaucouleurs+ 1991). *VizieR Online Data Catalog*, 7155:0, Feb. 1995.
- [8] J. C. Jacob, D. S. Katz, G. B. Berriman, J. Good, A. C. Laity, E. Deelman, C. Kesselman, G. Singh, M.-H. Su, T. A. Prince, and R. Williams. Montage: a grid portal and software toolkit for science-grade astronomical image mosaicking. *ArXiv e-prints*, May 2010.
- [9] D. Lang, D. W. Hogg, K. Mierle, M. Blanton, and S. Roweis. Astrometry.net: Blind Astrometric Calibration of Arbitrary Astronomical Images. *The Astronomical Journal, Volume 139, Issue 5, pp. 1782-1800 (2010)*, 139:1782–1800, May 2010. doi: 10.1088/0004-6256/139/5/1782.
- [10] T. Robitaille and E. Bressert. APLpy: Astronomical Plotting Library in Python, Aug. 2012. Astrophysics Source Code Library.
- [11] C. Stoughton, R. H. Lupton, M. Bernardi, M. R. Blanton,

S. Burles, F. J. Castander, A. J. Connolly, D. J. Eisenstein, J. A. Frieman, G. S. Hennessy, R. B. Hindsley, Ž. Ivezić, S. Kent, P. Z. Kunszt, B. C. Lee, A. Meiksin, J. A. Munn, H. J. Newberg, R. C. Nichol, T. Nicinski, J. R. Pier, G. T. Richards, M. W. Richmond, D. J. Schlegel, J. A. Smith, M. A. Strauss, M. SubbaRao, A. S. Szalay, A. R. Thakar, D. L. Tucker, D. E. Vanden Berk, B. Yanny, J. K. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, M. Bartelmann, S. Bastian, A. Bauer, E. Berman, H. Böhringer, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, L. Carey, M. A. Carr, B. Chen, D. Christian, P. L. Colestock, J. H. Crocker, I. Csabai, P. C. Czarapata, J. Dalcanton, A. F. Davidsen, J. E. Davis, W. Dehnen, S. Dodelson, M. Doi, T. Dombeck, M. Donahue, N. Ellman, B. R. Elms, M. L. Evans, L. Eyer, X. Fan, G. R. Federwitz, S. Friedman, M. Fukugita, R. Gal, B. Gillespie, K. Glazebrook, J. Gray, E. K. Grebel, B. Greenawalt, G. Greene, J. E. Gunn, E. de Haas, Z. Haiman, M. Haldeman, P. B. Hall, M. Hamabe, B. Hansen, F. H. Harris, H. Harris, M. Harvanek, S. L. Hawley, J. J. E. Hayes, T. M. Heckman, A. Helmi, A. Henden, C. J. Hogan, D. W. Hogg, D. J. Holmgren, J. Holtzman, C.-H. Huang, C. Hull, S.-I. Ichikawa, T. Ichikawa, D. E. Johnston, G. Kauffmann, R. S. J. Kim, T. Kimball, E. Kinney, M. Klaene, S. J. Kleinman, A. Klypin, G. R. Knapp, J. Korienek, J. Krolik, R. G. Kron, J. Krziesiński, D. Q. Lamb, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, B. MacKinnon, E. J. Manterny, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, B. McLean, K. Menou, A. Merelli, H. J. Mo, D. G. Monet, O. Nakamura, V. K. Narayanan, T. Nash, E. H. Neilsen, Jr., P. R. Newman, A. Nitta, M. Odenkirchen, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. S. Peterson, D. Petravick, A. Pope, R. Pordes, M. Postman, A. Prossapio, T. R. Quinn, R. Rechenmacher, C. H. Rivetta, H.-W. Rix, C. M. Rockosi, R. Rosner, K. Ruthmansdorfer, D. Sandford, D. P. Schneider, R. Scranton, M. Sekiguchi, G. Sergey, R. Sheth, K. Shimasaku, S. Smee, S. A. Snedden, A. Stebbins, C. Stubbs, I. Szapudi, P. Szkody, G. P. Szokoly, S. Tabachnik, Z. Tsvetanov, A. Uomoto, M. S. Vogeley, W. Voges, P. Waddell, R. Walterbos, S.-i. Wang, M. Watanabe, D. H. Weinberg, R. L. White, S. D. M. White, B. Wilhite, D. Wolfe, N. Yasuda, D. G. York, I. Zehavi, and W. Zheng. Sloan Digital Sky Survey: Early Data Release. 123:485–548, Jan. 2002. doi: 10.1086/324741.