Final Report to the Alfred P. Sloan Foundation

Grant Number: 2004-3-11

Grant Duration: July 2005 – July 2008

December 31, 2008

Prepared by

Dr. Richard G. Kron, Director
Dr. James E. Gunn, Project Scientist
Dr. David H. Weinberg, Project Spokesperson
William N. Boroski, Project Manager
Michael L. Evans, ARC Business Manager

Submitted by

Principal Investigator:

Dr. Richard G. Kron
Director
Sloan Digital Sky Survey
c/o The University of Chicago
Dept. of Astronomy & Astrophysics
5640 S. Ellis Ave.
Chicago, IL 60637
773-702-3335

Authorized Institutional Official:

Michael L. Evans
Business Manager
Astrophysical Research Consortium
c/o Office of Research
University of Washington
Gerberding Hall, Room G-80
Seattle, WA 98195-1202
206-685-7857

Final Report for the Sloan Digital Sky Survey-II

The Sloan Digital Sky Survey has been one of the most ambitious and most successful projects in the history of astronomy. In its first five years of operation (2000 - 2005), the SDSS obtained five-band imaging over 8,000 square degrees of the high-Galactic-latitude sky, detecting 215 million celestial objects. It obtained spectra of 675,000 galaxies, 90,000 quasars, and 215,000 stars, selected from 5,700 square degrees of this imaging. The data, fully calibrated and reduced, carefully checked for quality, and accessible through efficient data bases, have been publicly released in cumulative form, beginning with an early release of commissioning data and continuing with a series of annual data releases. Object catalogs, imaging data, and spectra are all available through the SDSS web site http://www.sdss.org, along with detailed documentation and powerful search tools.

Beginning in July 2005, the SDSS entered a new phase, SDSS-II, which continued through July 2008. SDSS-II used the SDSS telescope, camera, and spectrographs to carry out three distinct surveys to: 1) finish the spectroscopic coverage of the North Galactic Cap (the *Legacy* survey); 2) conduct an imaging and spectroscopic survey for stars in the Milky Way (the *Sloan Extension for Galactic Understanding and Exploration*, or *SEGUE,* survey); and 3) conduct a survey for Type Ia supernovae to measure the rate of expansion of the Universe (the *Supernova* survey). SDSS-II made all data publicly available via two data releases (DR6 in mid-2007 and DR7 at the end of October 2008).

These goals of SDSS-II were translated into the following practical terms: 1) create a contiguous area of sky, defined by a designated boundary, uniformly covered by imaging and spectroscopy using the same target-selection criteria as for SDSS; 2) obtain spectra of 240,000 stars with specific target-selection criteria towards 200 lines-of-sight sampling the whole sky visible from Apache Point Observatory; 3) obtain light curves of at least 180 Type Ia supernovae of sufficient quality to be useful for cosmological studies, and with sufficient speed of identification to enable timely spectroscopic follow-up. These goals are described in greater detail in the Project Execution Plan that was developed at the beginning of the SDSS-II.

Data Collection for SDSS-II: Goals and Results

<u>Legacy Survey</u>

Imaging

The imaging coverage of the North Galactic Cap had been almost completed as of the end of SDSS. The last Legacy imaging data were obtained in 2006-Q2. Cumulatively between SDSS and SDSS-II, we acquired 7646 square degrees in the North Galactic Cap, compared to the goal of 7700 square degrees.

We have recently created a data product that was one of the original scientific goals of the SDSS, namely an image of a stripe of sky, summed over many individual scans, that enables a relatively large area (250 square degrees) to be surveyed to relatively great faintness (factor of 6 fainter than a single scan). We call this data product the *co-add*; it and objects detected within it are available as part of the public data release.

Spectroscopy

SDSS-II Legacy acquired 524 new plates in the North Galactic Cap, which was sufficient to complete the spectroscopic coverage (accomplished in May 2008) in that region of the sky. These data yielded redshifts for an additional 250,000 objects spectroscopically classified as galaxies, and redshifts for 28,000 objects spectroscopically classified as quasars.

The total spectroscopic footprint for Legacy, SDSS+SDSS-II, is 8032 square degrees. Of this area, about 750 square degrees are in the South Galactic Cap.

Combining the results from both SDSS and SDSS-II, and including the three stripes in the South Galactic Cap, in total there are about 930,000 objects that are spectroscopically classified as galaxies and 120,000 objects spectroscopically classified as quasars.

SEGUE Survey

Imaging
SEGUE imaging was completed in January 2008, having achieved the goal of 3500 square degrees of new imaging. The scientific value depends on how these square degrees are distributed across the sky, and are not sensitive to the exact value of the total. The sky coverage map is given at http://www.sdss.org/segue/skycoverage.html. This map shows that we succeeded in sampling, with the planned sampling density, all of the sky visible from Apache Point Observatory (that is, regions of the sky having declination greater than -20 *deg*).

Spectroscopy
SEGUE collected 410 plates spread over 212 separate directions in the sky (*tiles*). These plates contain 237,000 spectra of stars based on SEGUE target selection (i.e., not including the 227,000 spectra of stars obtained by SDSS and SDSS-II Legacy). The originally stated goals for SEGUE were 400 plates (bright/faint pairs) covering 200 tiles, yielding 240,000 spectra.

The amount of data is only part of the accomplishment of SEGUE. SEGUE also developed a new software pipeline that yields a qualitative advance in information about the atmospheres of stars: temperature, pressure, and the abundance of heavy elements. The measurement of radial velocities was also greatly improved with the new pipeline.

Supernova Survey

The Supernova Survey carried out repeat imaging on Stripe 82 (the "southern equatorial" stripe) whenever such observations were possible during the interval September through November in 2005, 2006, and 2007. The 300-square-degree stripe had already been scanned more than 20 times in SDSS, and an additional 40 complete coverages of this area were obtained in SDSS-II. These scans were rapidly analyzed to identify potential supernovae so that likely Type Ia candidates could be followed up spectroscopically while they were still luminous. The necessary speed was enabled by announcing supernova discoveries promptly on the web and through Central Bureau Electronic Telegram circulars. The total count of spectroscopically confirmed Type Ia supernovae for the three-season survey is 498, far beyond what we had forecast. The data for this survey will include the measured brightness as a function of phase for each supernova, plus the redshift, spectral type, and other parameters.

Besides the Type Ia supernovae, we are planning to distribute photometric processing of all clearly detected transients (about 10,000 objects) for all three seasons, which is expected to become a major resource for archival study of the time-variable sky.

Summary Data Collection for SDSS-II

Data Release 5 was the final release for the first phase of SDSS (5 years), and Data Release 7 is the final release for both SDSS and SDSS-II (8 years total). Therefore, one might expect the volume of data in DR7 to be larger than in DR5 by a factor of about 8/5 = 1.6. This expectation is roughly fulfilled: some metrics are higher (number of stars with spectra is more than a factor of 2.0 larger because stars were deliberately targetted by the SEGUE survey), and some metrics are not as high (number of galaxies and quasars with spectra is up by a factor of 1.35, since some observing time was allocated to SEGUE and Supernova). The total number of spectra is close to 1.6 times as high in DR7 as it was in DR5. Thus not only did SDSS-II achieve its scientific goals, but it did so with comparable efficiency to what was accomplished in SDSS.

SDSS-II Budget and Spending

The SDSS-II project, as presented to the Sloan Foundation in our proposal of February 2004, included a total scope of work of $14.9M over the three years. This budget is repeated in the table below in the column called "2004 Proposal." Since that proposal was submitted, we raised cash from the participating institutions beyond what we had expected. The National Science Foundation urged us to develop a spending plan for this additional cash, that is, to increase the scope of work. Our revised budget featured a much more realistic amount for the Management Reserve. The new budget was written into the Project Execution Plan and approved by the Advisory Council in the Fall of 2005. This new budget is shown in the table in the column headed "Baseline." The final column, "Actual," shows what was actually spent over the three-year project in each of the categories.

In our 2005 "Baseline" plan we stated that unspent management reserve funds from the prior year would be used to fund other areas of the project, thus explaining the difference in spending by category between the "Baseline" budget and "Actual" spending.

The SDSS-II project ended under budget, with a cash surplus of $1.4 million from unspent management reserve and miscellaneous income. We will use half of the surplus funds, $700K, to set up long-term stewardship of the SDSS and SDSS-II data. The remaining funds, $720K, will be transferred to SDSS-III.

Our request in the 2004 proposal to the Sloan Foundation for support for SDSS-II was for $5.4M. This amount was awarded in Grant Number 2004-3-11. The payment schedule for this grant was:

3/2005 $2.0M
3/2006 $2.0M
3/2007 $1.4M

Funds were received by ARC as follows:

5/2005 $2.0M
12/2005 $0.7M
12/2006 $1.7M
11/2007 $1.0M

SDSS-II Budget Forecast
(Forecast as of 11/12/08)

Total Forecast - Organized by WBS (in $000s)

Calendar Year	2004 Proposal[1]	2005 Baseline[2]	2008 Actual[3]
1. Survey Management	1,888	1,660	1,641
1.1. ARC Administration	284	281	320
1.2. Office of the Director	285	263	254
1.3. Office of the Project Scientist	346	241	243
1.4. Office of the Project Manager	769	666	613
1.5. Office of the Scientific Spokesperson	204	209	211
ARC Support for Project Spokesperson	32	37	9
ARC Support for Collaboration Affairs	41	48	59
ARC Support for Young Astronomers Travel Fund	0	0	23
ARC Support for Public Affairs	67	35	79
Public Information Officer	64	89	41
Survey Management Sub-total	1,888	1,660	1,641
2. Survey Operations			
2.1. Observing Systems	3,223	2,372	2,101
Technical Support at APO	1,189	1,035	993
Off-mountain Technical Support	702	483	696
Plug Plate Production	488	575	209
ARC Support for Observing Systems	844	279	203
2.2. Observatory Operations	4,868	5,225	5,122
2.3. Data Processing	2,612	2,536	3,061
Data Processing Operations	1,874	1,612	2,102
Software and Data Processing Support	738	924	959
2.4. Data Distribution	1,209	1,655	2,170
Data Distribution Operations	842	1,343	1,705
Data Archive Development and Support	367	311	465
2.5. ARC Support for Survey Operations	369	150	36
Survey Operations Sub-total	12,281	11,938	12,490
3. New Development			
3.1 Segue Pipeline Development	0	407	435
3.2. Supernovae Software Development	0	308	179
3.3. Photometric Calibration	0	159	136
3.4. DA Upgrade	0	241	267
New Development Sub-total	0	1,115	1,016
4. ARC Corporate Support	182	177	192
5. Education & Public Outreach	0	215	119
6. Management Reserve	549	1,165	0
Total	14,900	16,270	15,458

1. "2004 Proposal" budget as presented in the February 2004 Sloan Foundation proposal
2. "Baseline" budget as approved at the November 2005 Advisory Board meeting
3. "Actual" final costs as presented at the November 2008 Advisory Board meeting.

Summary Scientific Achievements of SDSS and SDSS-II

The SDSS data have supported fundamental work across an extraordinary range of astronomical disciplines, including the large-scale structure of the Universe, the evolution and clustering of quasars, gravitational lensing, the properties of galaxies, the members of the Local Group, the structure and stellar populations of the Milky Way, stellar astrophysics, sub-stellar objects, and small bodies in the solar system. Recent analyses by Madrid & Macchetto (2006, BAAS, 38, 1286 http://adsabs.harvard.edu/abs/2006BAAS...38.1286M) rate the SDSS as the most productive astronomical observatory in 2003 and 2004 based on citations to high-impact papers published in those years, ranking ahead of the European Southern Observatory (which includes the Very Large Telescope), Hubble Space Telescope, the Wilkinson Microwave Anisotropy Probe (WMAP), and the Keck Observatory. (Analyses along similar lines, privately communicated to us by R. Williams, rate the SDSS as the second most productive observatory in 2005, behind WMAP, and the most productive in 2006.) Similar studies have been undertaken by Virginia Trimble (e.g. http://adsabs.harvard.edu/abs/2008AN....329..632T), with similar conclusions about the impact of SDSS data on the field. As of December 2008, searches using the NASA/ADS abstract service show that the SDSS has contributed to over 2150 refereed papers with more than 75,000 citations. Roughly half of these papers were written by astronomers outside the SDSS collaboration. Of the 100 most cited astronomical papers written since 2000 (< 0.1% of the total), 16 are SDSS papers. SDSS is second only to WMAP in the past four years in most-cited papers. With the broader scope of SDSS-II, the scientific impact of the SDSS seems guaranteed to increase with time.. These data have supported an enormous range of scientific investigations by astronomers around the world. A good popular introduction to SDSS science (especially the cosmological questions that provided much of the original motivation for the SDSS) can be found in the American Museum of Natural History Science Bulletin http://www.amnh.org/sciencebulletins/astro/f/sdss.20051208/.

Here we list some highlights among the many discoveries made by the SDSS. In some cases these discoveries emerged over the course of multiple investigations spread over several years; here they are listed roughly in chronological order of the first investigations. Since the final SDSS data sets are only now being analyzed for the first time, the list of scientific highlights continues to grow.

The discovery of the most distant quasars, powered by supermassive black holes in the early Universe. These systems are remarkable in themselves, and by illuminating their surroundings they allow us to probe the transition from a Universe filled with neutral hydrogen to an ionized Universe filled with protons and free electrons.

The discovery of large populations of sub-stellar objects. At the opposite extreme from quasars, the SDSS discovered large numbers of "failed" stars, not massive enough to ignite nuclear fusion reactions in their cores. Together, the SDSS and the Two-Micron All Sky Survey (2MASS) have produced the main data samples for systematic studies of sub-stellar objects.

Mapping extended mass distributions around galaxies with weak gravitational lensing. Galaxies are surrounded by "halos" of dark matter, but prior to the SDSS the extent of these halos was inferred only indirectly. Analyses of SDSS data used the subtle gravitational distortion of the shapes of background galaxies to demonstrate that dark-matter halos typically extend to several hundred kiloparsecs (close to a million light years).

Systematic characterization of the galaxy population. By providing high quality images, distances, and stellar masses and ages of hundreds of thousands of galaxies, the SDSS transformed the study of galaxy properties and the correlations among them into a precise statistical science, yielding powerful insights into the physical processes that govern galaxy formation.

The demonstration of ubiquitous substructure in the outer Milky Way. SDSS maps of the distribution of stars show that the stellar halo of the Milky Way is filled with complex substructure, probably a signature of its hierarchical buildup from smaller components. SDSS data show that the motions of stars are correlated with their chemical compositions, providing further evidence that today's stellar halo is a mixture of distinct families of stars. The SDSS has also identified thousands of the most chemically primitive stars in the Galaxy, which provide insight into the first epochs of cosmic star formation.

Demonstration of the common origin of dynamical asteroid families. Through the identification of moving objects in the imaging data, the SDSS has greatly increased the number of known asteroids and provided precisely measured colors for them. These data give an improved determination of the asteroid size distribution, and they show that the members of dynamical families also have distinctive colors, demonstrating their common origin.

Precision measurement of the luminosity distribution of quasars. The SDSS has precisely mapped the rise and fall of the population of quasars, the most luminous objects in the Universe, and thus the growth of the supermassive black holes that power them.

Precision measurements of large-scale clustering and cosmological constraints. The precise clustering measurements, made possible by the enormous size and high completeness of the SDSS galaxy maps, provide powerful constraints on the matter and energy contents of the Universe and on the nature and origin of the primordial fluctuations that seeded the growth of cosmic structure.

Precision measurement of early structure with the Lyman-alpha forest. Each SDSS quasar spectrum maps the distribution of absorbing hydrogen gas along its line-of-sight (the "Lyman-alpha forest"). The enormous numbers of SDSS quasars have made it possible to measure the clustering of the underlying dark matter distribution with high precision, at epochs when the Universe was only 10% to 25% of its present age.

Detailed characterization of small and intermediate scale clustering of galaxies. On scales of tens of thousands to tens of millions of light years, the SDSS galaxy maps allow precise measurements of the clustering of many different classes of galaxies (defined by luminosity, age, shape, and other properties). These in turn allow strong tests of galaxy formation theories and statistical determination of the relation between galaxies and dark matter halos.

Discovery of many new companions of the Milky Way and Andromeda. Careful analyses of the SDSS stellar maps have revealed nine new dwarf galaxy companions of the Milky Way, nearly equal to the number found in the previous 70 years, as well as two new companions of the Andromeda galaxy. These discoveries transform the status of the "missing satellite problem," which is one of the key challenges in understanding galaxy formation in a dark-matter-dominated Universe.

Discovery of stars escaping the Galaxy. A small number of stars in the Galactic halo have velocities close to, or exceeding, the escape speed. These stars have likely to have been ejected by violent gravitational encounters with the supermassive black hole at the Galactic Center. They provide information on the conditions at the Galactic center and on the shape, mass, and total extent of the Galaxy's dark matter halo.

Discovery of acoustic oscillation signatures in the clustering of galaxies. The SDSS achieved the first clear detection of this long-predicted cosmological signal, an effect of sound waves that travel in the hot early universe and imprint a characteristic scale on the distribution of galaxies. This discovery opens the door to a new method of cosmological measurement that is the key to BOSS, the largest survey of SDSS-III.

Measurements of the clustering of quasars over a wide range of cosmic time. The enormous size of the SDSS quasar catalog allows measurements of the large-scale structure of the Universe that reach to great distances and therefore probe early times. Recent measurements show that the most distant quasars cluster much more strongly than nearby systems, implying that they reside in massive concentrations of dark matter. The measurements are also being used to test models of the growth and fueling of the supermassive black holes that power the quasar population.

Half of these achievements were among the original design goals of the SDSS, but the other half were either entirely unanticipated or not expected to be nearly as exciting or powerful as they turned out to be.

SDSS and Career Development

The SDSS has proven an excellent environment for young scientists' career development because it has given them access to and a role in producing cutting edge data and because it has connected them to a network of astronomers at leading institutions around the globe. As anecdotal indication of the impact on young scientists' careers, we list below examples of individuals who started working on the SDSS as students or postdocs and have since assumed leading roles in the next generation of major astronomical surveys, including SDSS-III, the Dark Energy Survey, and the Large Synoptic Survey Telescope (LSST) project. All of those listed are now in faculty or equivalent level staff positions except for Padmanabhan, who holds a senior postdoctoral position at Lawrence Berkeley National Laboratory.

Jim Annis began working on the SDSS as a postdoc at Fermilab and is now one of the lead members of the Dark Energy Survey project.

Michael Blanton began working on the SDSS as a postdoc at Fermilab and is now the Data Coordinator for SDSS-III.

Daniel Eisenstein began working on the SDSS as a postdoc at Chicago and is now the Director of SDSS-III.

Zeljko Ivezic began working on the SDSS as a postdoc at Princeton and is now the System Scientist of LSST.

Huan Lin began working on the SDSS as a postdoc at Fermilab and is now one of the lead members of the Dark Energy Survey project.

Heidi Newberg began working on the SDSS as a postdoc at Fermilab and is now one of the most active U.S. participants in the Chinese-led LAMOST Project.

Robert Nichol began working on the SDSS as a postdoc at Chicago and is now one of the lead members in both the Dark Energy Survey and SDSS-III.

Connie Rockosi began working on the SDSS as an undergraduate at Princeton and is now the PI of the SEGUE-2 Survey of SDSS-III.

Nikhil Padmanabhan began working on the SDSS as a graduate student at Princeton University and is now a leading member of the BOSS team at Lawrence Berkeley Laboratory.

David Schlegel began working on the SDSS as a postdoc at Princeton and is now the PI of the BOSS Survey of SDSS-III.

Donald Schneider began working on the SDSS as a postdoc at the Institute for Advanced Study and is now the Survey Coordinator for SDSS-III.

Erin Sheldon worked on SDSS as a graduate student and is now contributing to the Dark Energy Survey at Brookhaven National Laboratory.

Michael Strauss began working on the SDSS as a postdoc at the Institute for Advanced Study and is now the Intersurvey Science Coordinator for SDSS-III and the Science Working Group Committees Chair and a member of the Science Council and Board of Directors for LSST.

David Weinberg began working on the SDSS as a postdoc at the Institute for Advanced Study and is now the Project Scientist for SDSS-III.

Brian Yanny began working on the SDSS as a postdoc at Fermilab and will contribute to the scientific analysis of the SEGUE-2 survey of SDSS-III.

In addition to these individuals who have assumed leadership positions in new major survey projects, there are many other young astronomers who have assumed faculty (or equivalent) positions by virtue of their association with SDSS. This summary is necessarily incomplete because "association with SDSS" is not well defined, but the group would certainly include: Andrew Connolly, Mamoru Doi, David Hogg, Michael Vogeley, Gordon Richards, Xiaohui Fan, Jeffrey Munn, and Christy Tremonti.

Long-Term Stewardship of the SDSS and SDSS-II Data

We are implementing the following plan for archiving and distributing the SDSS and SDSS-II data. Memoranda of Understanding have been written between ARC and four participating institutions: Fermilab, the University of Chicago Library, Johns Hopkins Sheridan Library, and Johns Hopkins Department of Physics and Astronomy. The MOU's cover a five-year duration and spell out the tasks that each of these institutions will undertake in that period related to archiving (ensuring the robust preservation of the data) or serving (making the data accessible to the public in efficient ways), or both. The number of partners ensures that there is substantial redundancy and is consistent with best practices in the field of archiving digital data. ARC, in turn, is providing funds to the institutions to enable them to carry out their tasks. These funds are derived from the surplus cash available at the end of SDSS-II. The surplus exists because we did not spend the budgeted Management Reserve.

In practice, the data available at http://www.sdss.org/dr7 will continue to be served without interruption in the following years, where the ARC Business Office takes over responsibility of oversight of the work, and the SDSS-II management structure is no longer needed.

Near the end of the current five-year agreement, in 2013, each of the institutions will evaluate its position to continue its work with a new agreement. Since preserving and serving data is built-in to the missions of digital libraries and Fermilab, it is expected that the institutions will agree to continue with no further cash provided by ARC. It is difficult to foresee what may develop in detail, but the existence of SDSS-III can only help, since SDSS-III depends on the availability of SDSS and SDSS-II data, and it will be natural to federate the data from all of the surveys.