Sloan Digital Sky Survey

2002 ANNUAL REPORT

23 Dec 02

TABLE OF CONTENTS

2002 Annual Report

1. Introduction

This report describes the accomplishments of the Sloan Digital Sky Survey team for the period January 1, 2002 to October 1, 2002. Two of the SDSS goals are the creation of a uniform, five-band, digital photometric map of nearly half of the Northern Galactic sky to faint magnitudes, and the creation of a large, homogeneous spectroscopic survey of objects selected from this photometric map. The extension of the photometric map to three stripes of the southern sky and a spectroscopic survey of objects selected from these three stripes are included in these goals. Collectively these maps and surveys constitute the major component of the Archive. The third, equally important, goal is the distribution of the SDSS data to the astronomical community.

The observation phase of the Sloan Digital Sky Survey began in April 2000 and since then we have accumulated more than 3600 square degrees of imaging data and more than 400,000 spectra of all types of objects. While we have not quite reached the halfway mark in the observation phase of the 5-year survey, which ends June 30, 2005, the data have already had an impact on the disciplines of astronomy, astrophysics and cosmology. The public can access the first increment of the Archive, the Early Data Release, and our catalog of distant quasars, including the most distant quasars, through the SDSS website. We plan to make our first major data release, DR1, during the first quarter of 2003 and it will be roughly six times larger than the EDR.

Two years ago our observing efficiency did not meet our requirements and the image quality did not consistently meet survey requirements. Throughout 2002, we exceeded our efficiency goals and consistently met the survey requirements for image quality whenever the atmospheric seeing met survey requirements. Unfortunately, the weather continues to be our biggest disappointment and it has prevented us from reaching our baseline goals for observing. Sections 2 and 3 of this report provide details of observing statistics and the performance of Observing Systems.

During 2002 we made substantial improvements to the pipelines and this successful effort is described in Section 4. The data processing team kept pace with the acquisition of new data and the incorporation of the much improved pipelines in the Factory as described in Section 5. Since pipeline development and data processing were such major efforts in 2002 this work is described in more detail than other efforts and the people who did the work are identified. In the future we will highlight the work in other facets of the SDSS and the people who do the work. Section 6, Science Highlights in 2002, shows convincingly that the SDSS will answer many of the unanswered questions that we posed when we began the Survey. Section 7, Outreach and Communication, describes our efforts to train the next generation of astronomers and to share our excitement with the public. Section 8 outlines our stewardship of the funds we received, Section 9 describes our financial plans for the year 2002, and Section 10 presents our outlook for progress in 2003.

2. Observing Statistics for 2002

We have imaged 44% of the baseline area for the Northern Survey and 99% of the baseline area for the Southern Survey. All of these data have been processed and calibrated and meet survey quality requirements. As shown in Figure 2.1, we have accumulated a total of 3419 square degrees of imaging data for the Northern Survey. During the period from January through September 2002, we obtained 919 square degrees of new imaging data, which is only 50% of the baseline goal of 1821 square degrees for the same period. The single largest obstacle to meeting the baseline goal was the weather.

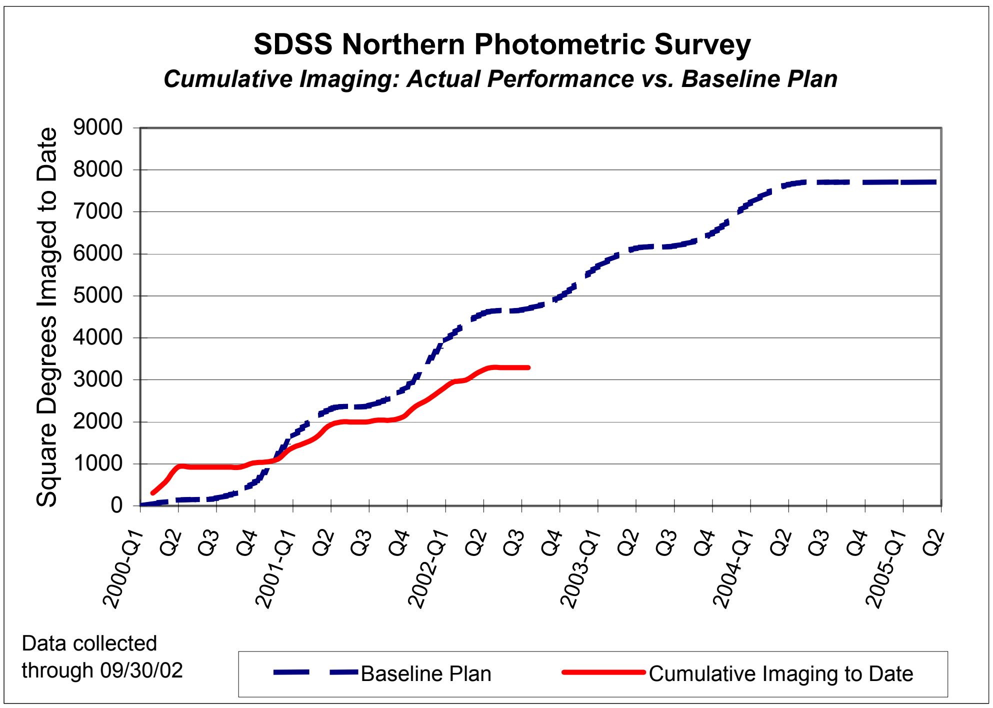

Figure 2.1. SDSS Northern Survey

The Southern Survey consists of three separate, non-overlapping, 2.5-degree wide stripes in the southern galactic hemisphere. The two outrigger stripes have a collective area of 475 square degrees and the central stripe, the Equatorial Stripe, has an area of 270 square degrees. As shown in Figure 2.2, we essentially reached the baseline goal for the Southern Survey at the end of 2001. In total, we obtained, processed, and calibrated 738 square degrees of imaging data on the three stripes. While this is short of the baseline goal of 745 square degrees, we have declared the Southern Survey complete because a significant investment in setup time would be required to obtain the last seven square degrees of missing data.

With the Southern Survey complete, efforts in 2002 focused on obtaining additional images of the Southern Equatorial Stripe. The cumulative area of imaging data obtained on the Southern Equatorial Stripe through September 2002 is shown in Figure 2.3.

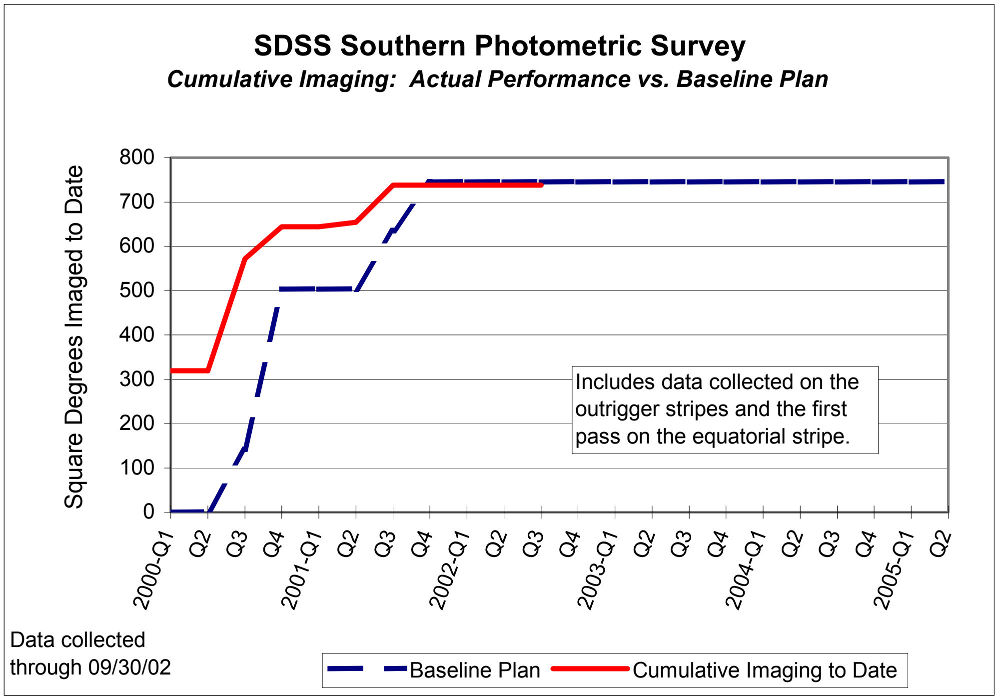

Figure 2.2. SDSS Southern Survey

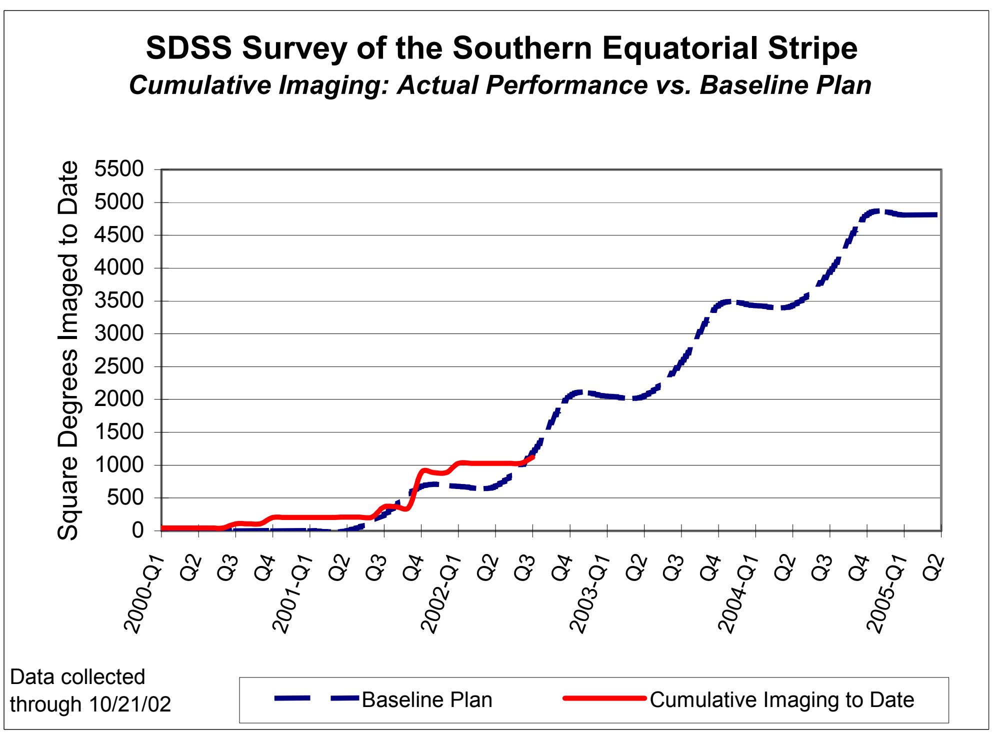

Figure 2.3. SDSS Survey of the Southern Equatorial Stripe

We also made steady progress on the spectroscopic surveys during 2002. Through the end of September 2002, our goal was to have observed 492 plates on the Northern Galactic Cap and 148 plates on the Southern Survey stripes. Cumulatively through the end of September, we observed 439 plates on the Northern Galactic Cap, 151 plates on the Southern Survey stripes, and 62 special plates. Since each plate yields 640 spectra, by now we have obtained slightly more than 417,200 spectra that meet survey requirements. Approximately 360,000 have both been classified with high confidence and have a high confidence redshift. The classifications of these spectra are shown in Table 2.1. The table does not include those spectra that were classified with lower confidence or the spectra that were dedicated to a special purpose, such as sky background.

Table 2.1. Classification of High Confidence Spectra

	Stars	Galaxies	Luminous Red Galaxies	QSOs	High-Z QSOs
Northern and Southern Surveys	40,240	229,253	31,475	38,707	2,454
Survey of Southern Equatorial Stripe	2,756	8,149	4,766	2,397	191
Total	42,996	237,402	36,241	41,104	2,645

We are behind the cumulative Northern Survey spectroscopic baseline goal by 53 plates. We were slightly behind the baseline at the start of 2002 and fell further behind in Q1 due to poor weather and Observing System problems. Although we recovered some of the difference with a strong second quarter performance, we nonetheless remain below the baseline goal through September. Table 2.2 shows the number of plates observed per quarter and cumulative performance against the baseline goal for the Northern Spectroscopic Survey We remain confident, however, that given the improvements that we have made in operational efficiency and our ability to exceed the baseline, as we did in Q2, we will catch up and meet the baseline goals for spectroscopy before the middle of 2005, when the Survey is scheduled to end.

Table 2.2. Progress in the Northern Spectroscopic Survey

2002 Period	Plates Observed	Baseline Goal	Cumulative Difference *(Plates Observed minus Baseline)*
As of Dec 31, 2001	234	263	-29
Jan-Mar	86	137	-80
Apr-Jun	110	82	-52
Jul-Sep	9	10	-53

Since we have observed 151 plates in the Southern Survey compared to the baseline goal of 148 plates, the spectroscopy on the Southern Survey stripe is essentially complete. In anticipation of this event, we issued a call for proposals to the collaboration in January 2002. We asked participants to submit proposals for additional spectroscopic programs that could be conducted on the Southern equatorial stripe area once the primary survey was completed. The proposal process was conducted in accordance with the Principles of Operation (PoO). 21 proposals were submitted, and a committee, appointed by the Director and chaired by the Project Scientist, reviewed the proposals and assigned time to the proposals that were compatible with the basic observing goals in the PoO. 17 proposals were approved, and an observing program was constructed so as to merge them together into a coherent program. The programs were judged on their ability to enhance the value of the science archive produced by the survey, enhance the science results being produced from that archive, and on their potential cost to the project. The proposers of each program were given the responsibility for producing the target lists and/or plate design files from the existing imaging data and object catalogs. These lists typically require a small amount of adjustment to make them conform to the survey standards, and they are then inserted into the normal cycle of plate drilling. The sum over all the programs is 334 plates and about one third of these have been designed and drilled.

Significant improvements in observing efficiency were made during the past year. Imaging efficiency, when measured relative to the baseline efficiency goals, improved from 80% in the fourth quarter of 2001, to an average of 94% through the first three quarters of 2002. During the February observing run, we deliberately chose to make short imaging scans and to aggressively push the limits of weather in order to acquire much-needed imaging data for spectroscopic target selection. While this strategy caused the poorer efficiency, it allowed us to build up an adequate inventory of spectroscopic plates for Q2. Spectroscopic efficiency improved in 2002, with efficiency increasing from 88% in the fourth quarter of 2001, to an average of 91% for the first three quarters of 2002. System uptime also improved from 91% in the fourth quarter of 2001, to an average of 96% through the first three quarters of 2002. Although the baseline goal for uptime is 90% and we are comfortably exceeding this goal, we are still striving for even higher uptime to ensure that all systems are fully functional whenever the weather is suitable for observing.

As noted previously, over the past year the weather has been the single largest impediment to meeting the baseline goals. As shown in Figure 2.4, there has been only one observing run in the past year when the amount of time that the weather was suitable for observing exceeded the baseline expectation. In addition, there were only a couple of months when the weather even approached the baseline expectation. When the weather does meet the baseline expectation, we are capable of strong data yields as demonstrated by the April observing run, in which we observed 62 spectroscopic plates and obtained nearly 40,000 spectra in a single month. We are ever hopeful that the atmospheric seeing will improve to at least the quality extrapolated from the experience of the 1990s.

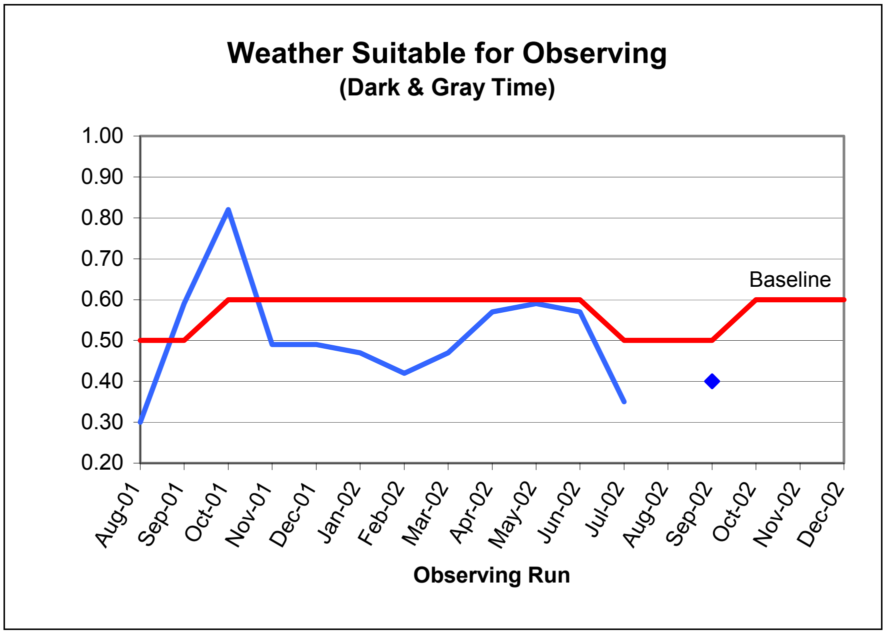

Figure 2.4. Fraction of Time Weather Suitable for Observing at APO

3. Performance of Observing Systems

3.1 Observing Systems

We began an aggressive effort in 2001 to improve the thermal environment of the telescope, since it was the cause of poor image quality in 2000 and early 2001. While most of the planned work was finished by the end of 2001, two projects were finished in 2002. One was the installation of louvers on the walls of the lower level of the telescope enclosure. They allow a significant flow of cold, ambient air to be pulled through the lower level and this airflow extracts heat generated by the equipment in the lower level. The second was the addition of a cooling system that extracted the heat generated by the flat field lamps and dumped it in the ground loop. Thermal images of the lamp bodies showed that they remained above the ambient temperature even thirty minutes after the lamps had been turned off. To prevent the "hot" lamp bodies from degrading the telescope seeing, plumbing was installed that connected the lamp bodies to the ground cooling loop, and this quickly cooled the lamps after they fired. We have finished all of our planned thermal improvements and they made a notable improvement in operational efficiency and data quality. The telescope reaches thermal equilibrium with ambient temperature within 30 minutes after the enclosure is removed. Prior to the improvement work, we lost up to one-half of a night waiting for the telescope to reach thermal equilibrium. In addition, the seeing measured by the 2.5m telescope is the same as the seeing measured by the APO 3.5m telescope. A year ago this was not the case.

The imaging camera and spectrographs performed well throughout the year, except during an unusually cold period in March when operations had to be suspended for several nights. The cold temperature prevented the spectrograph slit head doors and camera shutters from operating properly and seriously degraded the performance of several camera CCD chips and the camera T-bar latches. As a consequence of that experience, the SDSS engineering team made an inventory of the cold weather problems and assigned each one to various team members for follow-up action. The goal is to prevent the problems from recurring when the cold weather returns.

The software programs used by the Observers to operate the telescopes and instruments and acquire data have been a cause of system downtime and poor efficiency. Over the course of the year, software bugs were fixed and performance improvements were made. A critical bug that caused sporadic telescope axis halts was found and fixed early in the year. It would cause a problem only when a user process was controlling the telescope axes and a request for an interrupting output was made from that process. A second serious bug created a communication problem between the Spectroscopic Observers' Program (SOP) and the Telescope Control Computer (TCC) during the spectrophotometric calibration process. The source of the problem was identified and the required code changes made to correct the communication problem and improve the execution of control commands. Although a number of known minor bugs still exist, none are of a critical nature and the observers have devised work-arounds that keep these bugs from degrading operations.

In addition to fixing critical bugs, improvements were made to several software packages to improve performance and efficiency. Imaging setup time was reduced by improving the response of the data acquisition system (DAQ) to abrupt changes in ambient light level, which can occur at the beginning of a run when the shutters are opened at twilight or on "moony" nights. The response time was reduced from five minutes or more, to one to two minutes by rewriting the sky level package in the DAQ software. Additionally, all observing software routines were modified to record timestamps in UT format. The adoption of standard time formats made troubleshooting and analysis easier. Finally, it was decided at the May software planning meeting to limit code changes to only those necessary to address critical bugs or improve observing efficiency. A list of "final" changes was approved at the meeting and most of the changes were executed during the summer shutdown. The number of remaining planned software changes is now quite small. The most significant one will implement a new focus scheme for the spectrographs that will reduce setup time and thus improve efficiency. It is anticipated that restricting software changes will improve system stability and observing efficiency.

A number of engineering projects aimed at improving equipment protection, operations efficiency, and reliability were undertaken in 2002. In the event of an electrical power outage the Observers had to manually winch the telescope and enclosure into their respective stow positions. In bad weather this can be a strenuous and risky operation. A commercial gas-powered generator was installed during the summer that provides emergency power. Now the observers can safely stow the 2.5m telescope in the event of a site power outage. The actuator assemblies on the secondary mirror support system were upgraded to eliminate the failure mode that caused the loss of observing time last December and January. An LED illumination system for the imaging camera was designed, built, and delivered to APO. Its main purposes are to track CCD performance under illuminated conditions and to track changes in the flat fields. The systems provide data that has been incorporated into a new set of 'camera monthly checkout' scripts and the data is archived for use in calibration. Its importance was underscored this past year by the seemingly capricious changes in the flat-field response of the camera with time.

A new infrared all-sky camera system was designed and installed and is currently undergoing final commissioning. The new device, which operates in the 8-13 micron window, is about 30 times more sensitive than the old camera and can return frames with better signal-to-noise in 30 seconds than the old camera could return in 5 minutes. We replaced the cloud camera, which had limited documentation and no spares for critical components, because we were increasingly concerned with its long-term reliability and our ability to maintain it. Software for the new cloud camera prepares and displays the difference between the current frame and an average of up to 16 previous frames, and it computes a quantitative estimate of the "cloudiness."

We completed the design and initial installation of a new enclosure for the DIMM. The enclosure incorporates a pedestal, table, and pneumatically actuated cover. In addition, the DIMM telescope CCD was replaced with a CCD, with more and smaller pixels that will cover nearly twice as much sky. The new DIMM will also have two water-band filters that will make it possible to monitor the z-band water extinction. It will also provide automated coverage of many stars around the sky. The new cloud camera and DIMM will be completed by the end of 2002.

Finally, the lens on the guider was modified to prevent its movement during spectroscopic operations. Tests taken last May showed that the lens moved as the telescope moved and this caused a shift in image position, which degraded the efficiency of spectroscopic operation. A simple locking ring mechanism was designed and installed. Follow-up tests show that the locking ring reduced image movement by an order of magnitude, to 0.9 and 0.4 pixels in x and y respectively. Preliminary indications are that we may not have to replace the guide camera assembly, as earlier thought.

As in past years, the 2.5m telescope secondary mirror and the PT primary mirror were aluminized at Sunspot during the summer shutdown. The 2.5m primary mirror was aluminized at Kitt Peak at the end of September. The annual maintenance and refurbishment of the imager was carried out during the summer shutdown. Several camera problems were addressed, chief among them noise problems in g6, a seemingly successful fix to the intermittent u3 problem, relaxing tolerances to eliminate VDD errors seen during last winter's cold snap, and installation of new tertiary safety latches and a new pneumatic control/suspension system for the camera operations cart.

A formal program of scheduled maintenance tasks was begun. The new ODH system was repaired and calibrated. The first annual inspection of the instrument rotator bearing was made and the telescope cone azimuth bearing was cleaned and re-lubricated. Several years ago, we lost most of a dark run when this work had to be performed during a dark run. The 6-inch diameter hoses that route cables between the telescope fork and primary mirror cell, which are affectionately known as "elephant trunks", were modified.

The flexible plastic hoses began cracking during the March cold snap. New fittings were installed at both ends of each hose to allow free rotational motion and reduce stresses imposed at these points. It is anticipated that this improvement will eliminate the cracking problems seen last winter. Over the past year, we successfully devoted more time and resources to preventive maintenance activities, and we expect that this will pay off in terms of improved system uptime. During the coming year, we will continue the effort to formalize our preventive maintenance program and carefully assess our spares strategy, to further improve operational readiness.

Plug plate production and preparation operations ran smoothly throughout the year. From January through September, we drilled a total of 333 new plates. Target selection code was run on new imaging data and drill files were posted on the web in accordance with the monthly drilling schedule. The posted drill files were used to drill plates on a numerically controlled milling machine at the University of Washington (UW). After verifying plate quality through a set of QA measurements made on a coordinate measuring machine at UW, the plates were shipped to APO. The plates were received, inventoried in the on-line database, and prepared for spectroscopic observing according to instructions provided by the observers. Since the entire plug plate production process runs very smoothly, we are able to operate in just-in-time mode. This is necessary because we do not have enough imaging data to introduce a backlog of plates. Once we acquire enough imaging data, we will build some float into the plate production schedule to reduce our exposure to potential milling machine failures.

3.2 Observatory Support and Infrastructure

SDSS operations at APO ran smoothly and efficiently without unpleasant surprises. The SDSS Observers continued to concentrate on improving the efficiency of operations. They developed improved operational procedures and contributed to creation of better data-checking software as part of an increased emphasis on quality control. This effort will be given greater emphasis in the coming years as hardware and software problems require less attention from the observers. They also supported the development of the new cloud camera and the DIMM replacement. They also helped to troubleshoot other old and new systems. All site infrastructure support, such as the provision of liquid nitrogen, staff and visitor on-site housing, etc., was furnished as needed. The Observers operated and monitored the Photometric Telescope in its semi-robotic mode.

Although the Observers team was fully staffed during the observing season, the staff size was reduced by 0.5 FTE in the fall when the Fermilab staff member who had been spending 50% of his time supporting observing software and 50% as an observer was assigned to support the Tevatron Collider software. This reduction will be more than offset by hiring an eighth full-time observer. The additional position will help distribute observing responsibilities more evenly and will provide the team some flexibility in arranging their non-observing functional support projects. We further anticipate that as the observing systems reach maturity, the amount of bright time needed to test hardware and software changes will decline.

We were able to make substantive headway in addressing long-overdue site infrastructure improvements. A long-standing problem at APO was the poor quality of the office and work space for the SDSS engineering group and the SDSS Observers. After considering various options, we determined that replacing the two single-wide trailers, which had been on site since 1996, with two double-wide office trailers was the cost and time-effective solution. Installation of the new trailers occurred in 2002. The site preparation to accommodate the larger trailers was completed during the summer shut down, including the installation of plumbing and electrical service. In spite of the fact that problems were encountered during trailer delivery and set up, which took the vendor several months to rectify, we were given beneficial occupancy during the summer. As of the end of September 2002, both trailers meet requirements and are fully occupied.

A major facility milestone was the successful procurement and installation of a new “vice-over-IP” site-wide replacement telephone system. Under a cost-sharing agreement with the 3.5-meter telescope project, the site's 15-year old telephone system was replaced with a fiber-based system, which piggybacks on the site's existing local area network (LAN). The old system had become obsolete, could not be expanded, and frequently failed, particularly during lightning storms. The new system is relatively impervious to lightning and has features missing in the old system. Our early experience with the new system is that it works well.

Although there were several forest fires this year in the Sacramento Forest (one major), APO was not directly threatened. Negotiations continued with local agencies to initiate this year a fire fuel-reduction program near the site. Under a federal cost-sharing grant, APO has been selected as a pilot project to undertake tree thinning and ground fuels clearing in the down-slope up-wind direction around the observatory. We plan to treat roughly 8 acres of land before the next fire season.

Housing accommodations were provided at the observatory for visiting scientists, engineers, and other technical staff supporting SDSS operations. In addition, on-site housing was provided to the SDSS Observers when they were on shift. Table 3.1 shows on-site dormitory utilization over the period January through September 2002.

Table 3.1. APO Dormitory Utilization – January to September 2002

Average "traffic"	6.7 people/night
Percentage related to SDSS activities	69%
Percentage related to 3.5-m telescope operations	22%
Percentage related to other purposes (e.g., NMSU, other visitors)	9%

Over the course of the next year, in addition to supporting standard SDSS observing operations, we plan to train and fully integrate the eighth Observer, continue to improve operating efficiency, assimilate a visiting LAMOST scientist into SDSS operations as part of a joint collaborative effort between the SDSS and LAMOST projects, complete as-built documentation for observing systems, and assess the adequacy of our spare parts inventory.

4. Pipeline Development and Calibration - The Enabling Tools

Though the instruments and telescope of the SDSS were carefully designed and constructed to produce the data that has enabled the scientific results obtained to-date, it is the data reduction pipelines that have vastly facilitated their execution. Over the duration of the survey these pipelines have achieved a level of robustness and accuracy that produce results whose consistency and quality is limited only by the quality of the input data--surely the desired state for support software. The accurate photometric calibrations have played an equally important role in the quality of the SDSS publications. Science from quasar target selection to galaxy evolution was made possible by accurate photometry and pipelines. The three major pipelines, the Imaging Pipeline, the Spectroscopic Pipeline, and the Photometric Telescope Pipeline process the imaging, spectroscopic and photometric calibration data respectively.

The Imaging Pipeline is a chain of four pipelines that process the imaging data from the 2.5m CCD Camera. The data is first processed by SSC, the Serial Stamp Collector, which determines the sky level quartiles and selects postage stamps, which are images of small fields around bright stars. The outputs of SSC are then passed to PSP, the Postage Stamp Pipeline, which calculates the point spread function (PSF) as a function of position in the camera column and time along the scan and makes a preliminary photometric calibration of the selected stars using the appropriate outputs from the Photometric Pipeline. The outputs of PSP are then passed to the Astrometric Pipeline, ASTROM, which computes the transformation between pixel coordinates and astronomical coordinates. Finally, the outputs of the three preceding pipelines are passed to FRAMES, which identifies and classifies all of the objects that are above background. PSP and FRAMES constitute PHOTO. Lupton, Ivezic, Knapp, and Strauss produced SSC and PHOTO.

PHOTO does all of the imaging data reduction, removes all of the systematic errors introduced by the telescope and camera, and produces catalogs of deblended objects, which are classified by morphological type with photometry and shape parameters from several different schemes. Of very high importance in this work is the separation of stars and galaxies. The SDSS is a relatively shallow, very wide-angle survey and stars outnumber galaxies by a modest amount through most of the magnitude range of the survey. In order to accomplish this separation robustly and to do accurate photometry, it is necessary to characterize the point spread function and its variations with high accuracy: PHOTO uses a Karhunen-Loewe expansion of the PSF over the data, which allows for variation both across the scans due to optical variations and variation along the scans due to seeing. In addition, the burden of our goal of two percent photometry over the whole survey sky rests roughly equally on PHOTO and the external photometric calibrations.

Accurate photometry is a very high priority for the survey. Considerable attention is being paid to every detail in the CCD camera, the 2.5-m telescope, the Photometric Telescope, and the pipelines that process the data from these instruments. The SDSS photometry is connected to a set of standard stars through observations of these stars made with the USNO 40” telescope, which is equipped with SDSS filters. These observations define the (u’ g’, r’, i ‘, z’) system. Since the standard stars are between 8th and 13th magnitude they saturate the 2.5-m camera CCD’s when they are observed with the 2.5-m telescope. An intermediate step is used to transfer the (u’ g’, r’, i ‘, z’) system from the USNO observation to the 2.5-m observations and this is done with the Photometric Telescope (PT), a 20” telescope that uses the same type of CCD and the same type of SDSS filters used the 2.5-m telescope. The PT can observe both the standard stars and stars between 14th and 17th magnitude without saturating its CCD. The stars between 14th and 17th magnitude, which are observed by the 2.5-m system without saturation, are used to determine the 2.5-m PSF and the photometry of each 2.5-m field. The sky is observed with the PT in all five filters, concurrently with 2.5-m operations, and each field, called a patch, is processed with the Photometric Telescope pipeline, MTPIPE. The patches allow the comparison of the standard stars with the dimmer stars, which in turn enables the transfer of the standard system (u’ g’, r’, i ‘, z’) to the natural system of the 2.5-m telescope and CCD camera (u, g, r, i, z). Unfortunately, there are small differences between the three sets of filters, which are caused by the amount of water vapor in the respective filter environments. These variations complicate the photometric equation that defines the transfer between the USNO system and the natural system of the 2.5-m camera. The USNO-SDSS filters are mounted in air in Flagstaff and are stable. The 2.5m-SDSS filters are mounted in vacuum and are stable. The band-passes of the two sets of filters are slightly different. The original PT-SDSS filters were mounted in a box flushed with dry nitrogen and the water vapor content on the surface of the filters and the CCD was variable, which caused the width of the filter band-pass to vary. The filters for the PT were remade in 2001 using a different process that eliminated the sensitivity to water vapor. However, the band-passes of these filters are slightly different than the band-passes of the other two sets of filters. These differences were recognized in 2000, as noted in the 2001 Annual Report. A major effort was then made by Gunn, Finkbeiner, Hogg, and Schlegel that defined the photometric equations that connect the photometric systems defined by the USNO 40”, the PT and the 2.5m.

Since the strength of the electronic signal from the CCD’s is a nonlinear function of the photon flux that falls on each chip, the actual response of each CCD must be used in PHOTO to reach the goal of 2% photometry. To that end a JPG team, led by Mamoru Doi, designed and constructed a spectrophotometer, which was used to measure the response of all the CCD’s in the 2.5-m camera and the PT. During the last quarter of 2001 and the first quarter of 2002 the JPG team and the Observers measured the individual response functions and they are now used in PHOTO.

The pipeline that processes the data from the PT, MTPIPE, was developed largely by Kent and Tucker. It determines the extinction and the transformations that convert 2.5-m CCD counts to calibrated aperture magnitudes. The application of the calibration is done with a separate pipeline, NFCALIB, after the 2.5-m imaging data has been processed and stored in the Operations Database (OpDb). Lee and Lin produced NFCALIB and Munn and Yanny developed the OpDb.

The Astrometric pipeline, ASTROM, developed by Munn and Pier, determines the astrometric transformation by comparing the pixel coordinates of the brighter stars associated with each postage stamp to the astrometric coordinates of the same stars contained in an astrographic catalog.

A two-stage spectroscopic pipeline, SPECTRO, reduces the spectra obtained with the 2.5-m telescope. The first stage code, SPECTRO 2D, was developed largely by Schlegel and Burles. It produces calibrated one-dimensional spectra from the CCD images, removes instrumental signatures some of which (like the large-scale infrared scattering in the SITe detectors) are extremely difficult to manage, and applies spectrophotometric calibrations from simultaneous observations of F subdwarfs whose colors correspond to reliably known spectra. The spectra are then passed to the '1D' pipeline, SPECTRO 1D, which was developed largely by Bernardi, Subbarao, and Frieman. It classifies spectra, independently of PHOTO and determines their redshifts. It also determines the line strengths for common astrophysical spectral features.

4.1 Pipeline Development and Testing in 2002

There was a big push to deliver very high quality data for our first major data release, DR1. We chose to reduce these data with the best versions of the data processing software pipelines even though they were still under active development. We made substantial improvements to all three major pipelines, we reformulated the photometric and spectro-photometric calibrations, and we greatly expanded the number of checks that are carried out routinely on the data as part of quality assurance. These changes made quantitative improvements to the astrometric, photometric and spectroscopic outputs and they are described in the following paragraphs.

4.1.1 SSC and PHOTO (PSP and FRAMES)

The improved SSC and PHOTO v5_3 pipelines are the result of a two-year-long effort that the Princeton software group largely completed in July, when it was possible to routinely run production quantities of imaging data in the Factory. The principal changes made to SSC and PHOTO v5_3, relative to the previous versions of these pipelines, are:

- SSC now selects the brighter stars for postage stamps and makes a determination of the quartiles. In the past, these steps were performed in the Data Acquisition System and the outputs were used in SSC. Since these computations had to be done in real-time as part of the data acquisition process, the amount of memory and computing power available for the computations was very limited. They are now done in the pipeline environment, where much more memory is available and computation time is not a constraint. As a result, many more stars can be found in each field and this provides PSP with more information to characterize the PSF.

The results are much more accurate and reliable. In the past the limited computing resources and limited time occasionally lead to incorrect results.

- The order of processing steps was switched to allow PSP to follow SSC and thus to supply much more accurate information, including a preliminary photometric calibration of the postage stamp stars, to ASTROM. This calibration allowed the systematic errors due to asymmetry in stellar images to be corrected.

- The variations of the PSF with time and space are handled much more precisely in PSP and this led to superior photometry and improved star-galaxy separation, even in poor seeing. In addition, the outer portions of the PSF are now better characterized than they were in PHOTO v5_2.

- An improved deblender, which separates overlapping images, was incorporated into PHOTO v5_3. Previous versions of the code often shredded the images of galaxies into pieces; the new version behaves much better.

- The measured response function of each CCD was incorporated into FRAMES, thereby providing a much more accurate transformation between photon flux and electronic output signal than had been previously used.

- An improved cosmic-ray algorithm was implemented in PHOTO v5_3, which significantly reduced the false detection of rare objects such as brown dwarfs and high-redshift quasars.

- The calculations of measures of the shapes of galaxies, especially faint galaxies, were improved. This allows one to use the output of FRAMES directly to compute the weak gravitational lensing signal due to foreground galaxies and clusters, thus directly measuring the quantity of dark matter associated with these objects.

During the third quarter of 2002, a major effort was made to create accurate flatfields for the imaging camera. The method for flat-fielding the data was improved to take into account the presence of scattered light. This is necessary for accurate photometric calibration. Some of the specific problems that plagued processing at Fermilab in the past were eliminated.

Substantial additions were also made to the quality assurance tools that are run after each imaging run is processed. These come in two flavors: tests of internal consistency within a given batch of data, and the comparison of the outputs of each new run with already existing runs in the regions of overlap. These quality assurance plots are made after the output of FRAMES is loaded into the Operations Database. They allow the production team at a glance to determine that the data meet survey specifications.

4.1.2 ASTROM

The astrometric pipeline has been improved in a number of ways, including superior treatment of chromatic aberration. The incorporation of information from SSC and PSP, and the use of the UCAC catalog, a superior astrometric calibration catalog, has dramatically improved the astrometric solutions. The centroiding algorithms that was used in FRAMES, is now used in PSP and ASTROM. This also helped to make a more accurate and consistent astrometric solution. The rms position error was reduced from 80 mas to 45 mas and the previous systematic error of 50 mas in the PHOTO v5_2 reductions was eliminated.

4.1.3 MTPIPE and NFCALIB

The photometric calibration of the 2.5-m imaging data is done with a separate telescope, the 20-inch 'Photometric Telescope'. The calibration is complicated by the fact that because the effective band-passes of the filters used in the PT and the 2.5-m differ systematically. Our previous formalism for carrying out the photometric calibration took this into account in only an approximate, and not completely self-consistent way. A major effort was undertaken in 2002 to reformulate the photometric equations that are used in MTPIPE and NFCALIB, which applies the calibration to the 2.5-m data. This code is now in place and working, and has been shown to give superior results to the previous version.

A new version of NFCALIB v2_1, compatible with the new PHOTO v5_3 outputs, was cut in June 2002. These changes put our magnitude system on the 2.5-m telescope system. This will simplify the ultimate transformation of all SDSS photometry to a true AB system.

4.1.4. Target/Tile/Plate

The Target/Tile/Plate code did not require development although they were modified to accept data in the "PHOTO v5_3" format. We designed the calibration changes in MTPIPE and NFCALIB so that target selection algorithms would remain the same. This is important; we want to have the same criteria for the selection of galaxies and quasars for spectroscopy throughout the survey. Upon testing, we verified that these calibration changes were implemented correctly.

4.1.5 SPECTRO Development

The flat-fielding, bias subtraction, and handling of bad columns and pixels in the data in SPECTRO 2D was significantly improved. Sky subtraction was improved by introducing a gradient term in the sky brightness across a spectroscopic plate. The spectrophotometric flux calibration was improved, as was the correction for absorption lines from the Earth's atmosphere. There have been some problems with the calibration of the spectrophotometry, which are currently being addressed, and we expect the spectrophotometry to improve by a factor of about 3 (from about fifteen percent to about five) in the near future.

There have been upgrades to the continuum and line-fitting routines in the SPECTRO 1D pipeline. Improved stellar templates have greatly improved the classification of unusual types of stars. There are actually two completely independent 1D pipelines, an 'official' one and an unofficial one optimized for stellar classification. A careful comparison of the outputs of the two SPECTRO 1D output was carried out and it showed that both more than meet the requirements for redshift accuracy and spectra classification.

4.1.6 Scientific Testing

The science results obtained with the new pipelines and calibration procedures are being used in end-to-end tests of the entire system of pipelines. This has allowed many members of the Collaboration to become actively engaged in testing the data. The scientific testing is under the aegis of the testing group, which is led by Strauss. A number of bugs have been found and fixed. We expect to find more because the improved quality of the data and the improved performance of the pipelines will expose effects that were previously hidden.

Knapp and Strauss completed vetting the spectroscopic data that will be included in the first large data release (DR1) by comparing the outputs of the two different versions of the 1D codes. They obtained the astonishing result that in 98 percent of the reductions the outputs agreed as to type and redshift. Half of the 2 percent discrepant cases corresponded to spectra of very poor signal-to-noise in which a reliable classification is probably impossible, and most of the rest to rare (and sometimes bizarre) types of objects for which the two codes did not have overlapping template sets. We consider this level of performance to be extremely good and it clearly contributes to the accuracy of all the studies that make use of SDSS spectra.

The software performance is spectacular. It is one of our big successes for the year.

5. Data Processing and Distribution

The Fermilab Data Processing (DP) group carries out production data processing using programs developed by the participating institutions and scripts prepared by the DP group on computing platforms provided by Fermilab. The programs, scripts, and platforms for data processing create a highly automated environment that is called the Factory. Production scale samples of data are also processed with development versions of the pipelines in order to support the continuing development of the major pipelines. The output FITS files from the Factory are stored on disk and made available to the Collaboration for their science projects. The SDSS also plans to distribute the data to the collaboration and the public through two web-accessible systems because the sheer sizes of the FITS files make them difficult for many astronomers to use. One system, the Data Archive Server, is a set of web-based links that allow astronomers to efficiently extract subsets of the FITS files for subsequent analysis. The other system, the Catalog Archive Server (CAS), is a database that was designed to enable astronomers to make catalogs quickly and efficiently. The CAS and the tools to access and exploit the CAS databases are being developed at Johns Hopkins University. An early development version of the CAS was made available to the public with the release of the data obtained prior to April 2000. This release, the Early Data Release (EDR), was issued in June 2001. One of the SDSS goals for 2003 is the completion of these systems in time for the first major release of SDSS data, Data Release 1 (DR1). We had scheduled to issue DR1 in January 2003 and now we anticipate releasing it with the DAS in the first quarter of 2003.

One of the primary SDSS goals for 2002 was the complete processing of the DR1 data set with new versions of all of the pipelines. The DR1 data set consists of 64 survey quality runs obtained prior to July 1, 2001 that were used to target objects for spectroscopy (target runs) and the 63 best runs of survey quality image data (best runs) obtained prior to July 1, 2001 that cover the same area of sky covered by the target runs. In one instance, data obtained for the Southern Equatorial strip after July 1, 2001 was substituted for earlier data, when the best runs for that area of sky that did not meet survey requirements. The 63 best runs cover 2659 sq deg and contain 77,070 fields. A field is .0333 sq deg. The 64 target runs cover 2700 sq deg and contain 78,258 fields. This area of sky was partially covered by 291 spectroscopic plates yielding 186,240 spectra. The full DR1 data set requires nearly 3TB of disk storage.

5.1 Data Processing Operations

The morning after data are obtained on the mountain, imaging, spectroscopic and PT data are written to tape. The APO Observers' logs and the astrometric data, the GANGS files, are transmitted to Fermilab over the Internet since the data volume is quite manageable. The spectroscopic and PT data sets are also transferred to Fermilab via the Internet through automatic scripts and processing of these data streams can begin a few hours after observations are completed at APO. Since the imaging data sets are much too large to use the Internet link between APO and Fermilab, the tapes are shipped via overnight delivery to Fermilab. Processing typically begins about 24 hours after the tapes are packaged for shipment. All of the data tapes are eventually shipped to Fermilab where they are placed in long-term storage.

Once the imaging tapes arrive, the logs are examined and lists of runs are selected for processing, with start and end field ranges selected based on the APO night logs. Data acquired under obviously bad observing conditions, such as out of focus frames and non-photometric scan segments, are not processed. The data selected for processing are spooled from the imaging tapes onto disk as the first link in the chain of processing pipelines. The entire chain is automated and quality check scripts are run at the conclusion of each link. If the check passes, the Factory automatically chains to the next pipeline and processing continues. Each link can be processed on different computing platforms and typically the processing of data through different links of different runs is in progress simultaneously. In addition, parallel processing is used within each link whenever it will significantly reduce the processing time. For example, up to 120 farm nodes can be used to independently process each of the six camera columns of a typical FRAMES run of 400 fields (2400 fields in 5 colors) in less than 4 hours. Processing stops when a quality check fails and this permits intervention to determine the nature of the problem and whether processing should continue. In the event that it cannot, either some of the data does not meet survey quality or the code or scripts need to be modified before continuing. In the former case the fields that cannot be processed are set aside and processing resumes. In the latter case a problem report is filed and processing is resumed once the problem is fixed.

The complete chain of imaging pipelines in the Factory consists of the following links: Spool, SSC, PSP, ASTROM, FRAMES, OpDb-stuffing, OpDb-merge, NFCALIB, OpDb-setQA, OpDb-checkQA, zoom-generation, and calibrated-run-export. There are several points along this chain where backup processes are spawned in the background that package and copy processed data to the Enstore tape robot. The raw data is also backed up at the start of the processing sequence. Web pages that describe the status of processing for every run and that provide links to the quality plots are automatically updated.

Unless there is a problem, new imaging data from the mountain can be processed thru the 'calibrated-run-export pipeline stage' into flat FITS tsObj files within 72 hours of receipt of the tapes from the mountain. At this stage the processed data are available for scientific use by the Collaboration. Historically problems that stop processing occur in about 15% of the runs. These runs take much longer to process since in some instances they require the developer to correct a pipeline bug.

Spectroscopic processing is handled entirely by automated scripts that look for the data on the disks at APO, pull them to Fermilab and subsequently feed them to the spectro pipelines (SPECTRO 2D and SPECTRO 1D). Since the outputs from SPECTRO 2D and 1D are not very large, they are placed on disk that can be accessed by the Collaboration. The processing is done on the farm nodes using high level processing scripts that were developed by the DP group. Spectroscopic data and PT data are generally processed within 24 hours of when they are obtained on the mountain.

Automated scripts developed and maintained by the DP group control the PT data processing. The group is also responsible for day-to-day management of PT data processing and oversight of quality control.

Adelman selects the imaging data that is suitable for processing based on the nightly observing logs, with scientific oversight by Steve Kent. Stoughton, Adelman and Hendry monitor the progress of the imaging data through the chain of imaging pipelines. Adelman, with scientific oversight from Yanny, Kent, Lin and Jester, reviews the quality plots produced at the end of each link in the pipeline chain. Inkmann and Adelman update the web pages that describe the current state of the processing of each imaging run. Tucker selects the PT data for processing and monitors the progress of the processing through MTPIPE. Once the processing has produced a set of patches, he performs QA on the process. Lin, Kent and Tucker review the quality of the patches that are available to calibrate the imaging data and select the best set of patches for calibration of a particular set of runs. Lin carries out the calibration of the imaging data with NFCALIB. When this step is complete he performs target selection with TARGET and then produces the spectroscopic plate drilling files each month. Kuropatkine and Burles developed the high level scripts to process the spectroscopic data on the Fermilab computing farms and

Kuropatkine monitors the progress of spectroscopic processing. Science oversight of the spectro quality is to be handled by Annis.

Stoughton oversees the overall production system, D. Yocum maintains the (non-farms) Linux hardware and software, while L. Giachetti with support from Computing Division staff maintains the OSF1 and IRIX machines.

5.1.1 Image Data Processing

The DP group processed all imaging data obtained prior to the August 2002 dark run with PHOTO v5_2, the production version prior to that time, and distributed them to the Collaboration. The DP group also processed selected runs of the DR1 data set with PHOTO v5_3, which was under development, during the first six months of 2002. The developers used the PHOTO v5_3 outputs to test fixes of bugs that had been found in PHOTO v5_2 and to evaluate the new features that they were adding. In July, when the development of PHOTO v5_3 was essentially completed, it was possible to begin processing production size samples of data that were representative of the entire DR1 data set. These samples were used to determine the flat fields for the entire DR1 data sample. After carefully reviewing the flat fields obtained in this manner, the developers determined that it was necessary to create six different sets of flat fields in order to accommodate changes that were made to the camera vacuum system between September of 1998 and August 2002. Production processing of the DR1 data with PHOTO v5_3 was begun in late August when the first set of flat fields was completed. The flat fields and production processing of the DR1 data set were both completed at the end of September. Throughout the entire reporting period, the Collaboration was given access to data processed with PHOTO v5_3 and many collaborators were able to evaluate the performance of PHOTO v5_3 with those outputs. The testing exposed some problems that were subsequently fixed, and this led to important improvements in PHOTO.

The DP group began processing new imaging data with PHOTO v5_3 starting with the August 2002 dark run, and they are now gradually reprocessing the imaging data taken between July 1, 2001 and the end of the July 2002 dark run with this version when time permits. They expect to complete the reprocessing early in 2003, provided no further changes need to be made to PHOTO. Priority within the DP group is given first to processing new data and then to their support of the effort for the creation of the DAS and CAS for DR1. The latter continues to require significant effort from all DP members.

5.1.2 PT Data Processing

The offline processing of the PT data with MTPIPE is current and up-to-date. All survey-quality PT data from late January 2000 through April 2002 have been run through MTPIPE v8_0, which makes use of new photometric equations. The outputs have been used to calibrate the DR1 imaging data set with the new photometric equations. PT data obtained before February 2000 do not consistently meet survey quality and, except in a few special circumstances, they have not been re-reduced with MTPIPE v8_0. Data obtained since April 2002 have been fully processed with MTPIPE v8_0, for the calibration of imaging beyond DR1.

5.1.3 NFCALIB Processing

All new imaging data processed with PHOTO v5_3, as well as all older data reprocessed through PHOTO v5_3, are now routinely calibrated using NFCALIB v2_1. In particular, all DR1 imaging data have already been successfully recalibrated.

5.1.4 Target/Tile/Plate

Target selection and tiling were run for survey areas in stripes 9, 12-13, and 30-36, resulting in a total of 220 new main survey plates. In addition, 113 special plates were designed for a number of science programs constituting the new Southern Equatorial Survey on stripe 82. A grand total of 333 plates were designed and plate files were provided on time to the UW machine shop for every scheduled plate-drilling run during this period.

5.1.5 Spectroscopic Data Processing

All spectroscopic data taken prior to the August 2002 dark run have been reprocessed with the latest version of Spectro (rerun19) and new data, including data taken during the August dark run are being processed with latest version of Spectro. Spectroscopic data processing is current with its acquisition as of the beginning of October. As the reporting period ended a newer version of Spectro (rerun21), which improved spectro-photometry, was being tested and the DP group plans to use it for production processing as soon as it is available.

5.2 Data Distribution

The SDSS will distribute the data to the astronomical community through two types of data servers, the Data Archive Server (DAS) and the Catalog Archive Server (CAS). The DAS is a web interface to the FITS files of processed imaging data and spectroscopic data. It provides a set of tools to extract selected subsets of data from these files. It also provides links to the documentation that describes the properties of the processed data and how a user can obtain the data.

The CAS will provide the user with much more. It will support very complex queries that will enable the user to create catalogs of objects with their attributes from the SDSS data. It has been designed so that the creation of catalogs can be done quickly and without a big learning curve. The Johns Hopkins University (JHU) team of Szalay, Kunszt, and Thakar began the development of the SX database, a CAS that uses a commercial object-oriented database, and a query tool sdssQT, which allows a user to build catalogs of objects, more than five years ago. The sdssQT access method derives its power by keeping indices on the data for fast lookup and by making all the data available through a uniform portal. A great deal of effort had to be invested to adapt the object-oriented database to the SDSS data set. By late 2001 the development of the SX was sufficiently advanced to allow the collaboration to begin using the SX extensively. Unfortunately we learned that it could not meet our requirements without a major redesign. This was largely due to the reluctance of the vendor to make the changes in its product that would allow us to meet our goals. We had intended to use the SX to distribute the DR1 data set to the astronomical community in January 2003 and these difficulties prevented that from happening. The SX was updated with imaging and spectroscopic data obtained prior to July 2002 and the DP group and the JHU team will continue to support the SX, without further data updates, until the new CAS can replace it.

In parallel with the development of the SX, Szalay and Gray and a several of their colleagues at JHU and the Microsoft Bay Area Research Center (BARC) had built the SkyServer to give students and teachers access to the public SDSS data. They had initiated the SkyServer, which uses a Microsoft SQL database and server, as part of an outreach and education effort. Although the support for the Sky Server came from sources other than SDSS funding, it was included with the June 2001 Early Data Release (EDR). It was very successful and very popular with its intended audience and we also learned that it was a useful and painless way for professional astronomers to access the SDSS data for their research. In January 2002, we chose to stop further development of the SX since the Sky Server had many of the features of an ideal CAS and it appeared that it would be a relatively easy and quick transition. While we had reservations about such a dramatic change at this stage in the development of the CAS, we were forced to take this path when the difficulties of implementing the object-oriented version of the CAS became insurmountable.

The data moves through several stages of processing before it ends up in the CAS SQL database and can be queried by a user. First the CSV file maker converts each processed data file corresponding to a single imaging run, a single spectroscopic plate, or single tiling file from FITS format to ASCII format. Next the files and the directories are validated to verify that all components are present and that no errors have occurred during the conversion. Next the Load Server loads the CSV files for imaging (with the masks), tiling, and the spectroscopy into a small SQL "Task" database.

When the loading for this area of sky is complete, several tests are performed including a comparison of the number of lines in each file that was loaded to the actual number of objects in the database tables. In the next step, Validation, about a hundred different integrity tests are performed, like testing the uniqueness of a set of key fields, and comparing the actual cardinalities of tables, using various checksums, carried along from the export, like the numbers of objects per field, the number of fields per segment. Once these tests are successfully completed, a backup of the Task database is created and then the Task database is "Published" into its final destination. Following the Publish step, we delete the Task database in order to free up disk space for the next load of CSV files. Once a batch of Tasks is loaded, the Finish step, which builds all the necessary indices, computes neighbor tables for fast spatial access, and builds links between the Target, Spectro and Best datasets, is performed.

All through this process, we perform a four level logging and monitoring process. The loading of every task database creates at least a hundred phase messages, which can be used to monitor if there was any error or warning throughout the loading process. As a result the load history of any single object can be tracked, even a few years later.

The whole loading framework has been built in such a fashion that any part of it can run in parallel, across multiple servers if there is a need for very high ("campaign") throughput and it can also use a single server while in incremental ("steady-state") mode. Since the DR1 data set amounts to nearly one TB of data when stored in the CAS (the merged SQL database), there is a premium to minimize the time for each step in the creation of CAS, and the developers spent considerable time in building a robust workflow environment that is easy to run and maintain.

The DR1 CAS was designed from the beginning to be part of a full-fledged production facility that will allow the CAS to be updated automatically when new data is released. The developers spent a substantial amount of time developing a process that would find all of the problems in the data that would create errors during loading. As noted earlier there are actually three different data sets, the imaging files (with their masks), the tiling files, and the spectroscopic files and the boundaries of these files on the sky must be compatible and the objects that were selected for tiling and spectroscopy must be recognizable as the same objects. Moreover there are two flavors of the imaging files, the TARGET files and the BEST files, and there can be small differences in the location and number of objects and the nearest neighbors of the objects in the two files.

During the creation of the EDR many small problems were encountered that were caused by small discrepancies in the definition of the boundaries of these data sets and inconsistencies in the content of these data sets. It was possible to correct these problems by patching them by hand. The enormous size of the DR1 data sets made this impractical, since the data loading and validation had to be fully automated. A very serious multi-tier data validation and verification process was initiated to find and correct the discrepancies and errors that resulted in misalignment of the data sets. Until this was done the different data sets did not quite fit together. To make everything fit it was necessary to make small changes in the data model and then re-export parts of the data until all the pieces fit. It took considerably longer to develop and complete this process than anticipated. While it was a time consuming, iterative process for both the developers and the DP group, it led to considerable improvements in the integration of the data into the SQL database. This extensive "scrubbing" of the data has produced an excellent database that should allow us to go forward in the future with considerably less difficulty.

5.2.1 Status of the Data Archive Server for DR1

A serviceable version of the Data Archive Server (DAS) was completed and made available to the Collaboration. It consists of the DR1 FITS files and the tools that allow an astronomer to access the FITS files. It also includes the documentation that is necessary to use the files. The documentation describing the DR1 data set and the tools are presented through the SDSS website. After the SDSS collaboration has used the DAS for three months we expect that all serious problems will have been exposed and fixed.

The purpose of the three-month period is to allow a full evaluation of the links to the files of processed data, the access tools, and above all the quality of the DR1 data set. Our experience with the EDR revealed that without an extensive evaluation the data, the access tools, and the database could contain flaws, which would present serious problems to the astronomical community. We now plan to release the DAS to the public in the first quarter of 2003, several months later than we had planned four years ago when we created the data release schedule.

The DR1 FITS files have also been converted to ASCII format and are stored in a simple SQL database. A MySQL query tool has been developed that allows the user to select classes of objects with their attributes from the simple database. This query tool has limited functionality and is not intended to efficiently support complex queries. We anticipate that a user will extract a subset of the data from the DAS and export it to disk at their home institution. While the time that is needed to assemble a catalog with the simple SQL interface can be quite long, it should be a useful interim product. It should allow astronomers to create catalogues of SDSS objects based on a few properties like magnitude and color. When the CAS has been completed and thoroughly tested by the Collaboration, it will replace the interim SQL interface.

5.2.2 Status of the Catalog Archive Server for DR1

Progress on the new CAS has been excellent during the past year. All of the code and scripts required to implement the multiple steps in the creation of the CAS have been coded and tested by the developers at JHU and BARC. While the developers were able to salvage a substantial amount of the code from the SX CAS development, they still had to write over 20,000 lines of SQL code fro the DR1 CAS. A ten percent sample of the DR1 data set has been successfully loaded into the development CAS system at JHU and is being evaluated by the testing group. Since the hardware and operating system at JHU are capable of supporting a complete DR1 CAS, the development CAS will be loaded with the DR1 data set and tested at JHU before installing the production CAS at Fermilab.

We expect to transfer the full DR1 data set to JHU in December, and then load it into the CAS. At that time the entire Collaboration will be given access to the development DR1 CAS. In parallel with this effort the DP group will install and configure the hardware required for the public CAS at Fermilab. Since the hardware procurement has been placed, delivery of equipment should be complete by the first week of January (2003) and the system should be ready for the installation of the production CAS in early January. Once the full DR1 CAS has been successfully used by the Collaboration it will be installed at Fermilab. While the Fermilab computing environment delayed the start of the installation of the CAS at Fermilab, these problems appear to be solved. Moreover the creation of CSV files has also been carried out successfully at Fermilab. After the DP group demonstrates that the CAS installation at Fermilab is complete, has achieved stable operation and is bug-free, the Collaboration will be given access to the production CAS. Currently we expect to complete the loading of the production CAS with the entire DR1 data set during the first quarter of 2002. We expect that after the Collaboration has used the production CAS for three months any remaining problems will be exposed and fixed.

We expect to make the production CAS available to the public during the second quarter of 2003.

We did not foresee the difficulties of creating and testing the CSV files. The culling of bad fields from the imaging data and the subsequent determination of the spectroscopic tiles that matched the fields of imaging data was much more labor intensive than we expected. This was due to small, extremely subtle oversights in the data model. At the same time it is worth mentioning, that during the development of the database loading tools, we created a process that systematically identified the errors and then eliminated them one by one. In addition there was case when a file was corrupted when a disk failure and the RAID 5 system rebuilt the file with errors. We can now deal with this type of problem as well.

Recently we discovered that the network security requirements imposed by Fermilab in order to meet DOE requirements could affect the safe operation the CAS. Tests of the hardware and operating system that we plan to use for the CAS will be made in November. This will allow us to determine whether there is degradation of security and if there is, how it might be fixed. Because of 9/11 many changes have been made and continue to be made to the computer security environment at Fermilab. While none of these changes have created a fatal problem, the time that developers had to take to adapt to the new environment delayed the creation of a working CAS. While the discovery of these problems delayed the selection of the hardware for the CAS distribution platform, the hardware needed to carry out these steps and provide the disk storage for the staged databases and the merged database has now been defined.

5.2.3 Status of Data Distribution

The data obtained prior to the August 2002 dark run was processed with production versions of the pipeline software and made available to the Collaboration in the form of tsObj files. The DR1 data set, essentially all of the data obtained prior to July 2001 was reprocessed with the newest versions of the pipelines and made available to the Collaboration at the end of September in the form of tsObj files. As noted earlier these files are difficult for the typical astronomer to use. The SX database, which allows much easier access to the data, was made available to the collaboration throughout 2002. While it would crash from time to time, it gradually became reliable during the first half of 2002. The SX database is accessible through the SDSS website and is typically processing several queries at any given moment in time. Access to the SX is limited to the SDSS Collaboration. The SX database was kept current by adding new imaging data processed with PHOTO v5_2 and new spectroscopic data processed with rerun 19 until July. While JHU and the DP stopped adding new data to the SX in June and stopped improving its query tool sdssQT at the same time, they will continue to maintain the SX until the new CAS is fully operational.

The Early Data Release, EDR, continues to be distributed to the public through the SkyServer interface, which can be accessed through the SDSS website.

5.2.4 SDSS Help Desk

A help desk is being established by the DP group at Fermilab in order to respond to problems in the DAS found by the collaboration. We anticipate that this will test all aspects of the DAS, including the process for supporting the astronomical community. The responsibilities of the help desk will be expanded to include support for the CAS, once the collaboration has been given access to the production version of the CAS. It was necessary to do this when the STScI was not funded to support the public release of DR1.

5.3 Hardware Additions to the Factory

The hardware, which is used to support data processing and store the output in an accessible form, was substantially augmented in 2002 by constructing multi-Terabyte (TB) disk farms. Nearly 15 TB of processed data is available to the Collaboration on spinning disk. In addition, all code and data have been backed up either on Enstore, the tape robot system used by Fermilab, or DLT tapes located at the Feynman Computing Center. A limited amount of disk storage is used to backup actively used code and processed data. During the past year a major effort was made by the DP group to develop a reliable and robust back up of all SDSS files. The adaptation of Enstore to the needs of the SDSS was an important part of that effort. The use of Enstore is particularly important since the demand for disk space grew very rapidly during 2002 because it is necessary to make several versions of the processed data available to the Collaboration for testing and science. Ultimately, Enstore will be used to store image data processed with versions of pipelines that precede the current production versions and that are not in active use. In addition to substantial increases in disk storage, the computing power has been increased to allow a more rapid turn around of imaging and spectroscopic data.

6. Science with The SDSS: A Summary of Accomplishments to Date and Goals For The Rest of the Survey

This report is a bit broader than the science specifically accomplished in the past year, but we are approximately a third of the way in, and it seemed reasonable to summarize our overall progress toward the scientific goals we set ourselves at the onset of the project. The scientific goals of the survey have been laid out in the Project Book, which has formed the basis of essentially all of the funding proposals from the SDSS since its beginning. The Principles of Operation calls these major goals the `Core Science' of the survey. They are captured to a large extent in the list of Key Projects at

http://www.sdss.org/collaboration/key_projects.html

The key projects were compiled in 1997, though the list is substantially the same as the primary science goals contained in the Project Book, which is much older. It is important to recognize that this list is essentially as 'fresh' today as it was when it was compiled; certainly the goals looked toward the unique capabilities of the SDSS and it is perhaps not surprising that the problems have not been solved in other ways. It is perhaps surprising that the important questions have not changed very much, given the rapid development of the field. In many cases, especially in the very active area of large-scale structure, the questions have become much more focused, but their essential nature remains.

Let us first address these 'core' goals and progress toward meeting them in a general way. We must also consider that a very large amount of other excellent science has been done, some less central to the original conception of the survey (for example, the very important solar system results), some resulting from even better performance than anticipated (e.g. the weak lensing results), and yet others which take advantage of unanticipated astrophysical quirks (the very successful search for methane brown dwarfs.)

6.1. Large-Scale Structure

The survey was designed around and largely motivated by large-scale structure investigations, and we now have enough spectra and sky coverage to attack these questions meaningfully. We have to date about 220,000 unique main sample galaxy spectra, (out of a total of 411,000 acceptable spectra as of October 2002), which is slightly larger than the 2dF sample. Work on the current sample was reported in a series of papers at a special session at the June 2002 AAS meeting at Albuquerque. For 2D statistics, work on photometric redshifts with the now well-understood SDSS photometric system has essentially reached fruition, and results were presented on the two-point correlation function using photometric redshifts to isolate shells in distance by the LSS group led by Szalay. We expect eventually that the 2D investigations with photometric redshifts will be more powerful than the full 3D investigations because of the much larger number of objects and greater depth of the photometric sample. The 3D covariance function and power spectrum work is much more mature at this point, however, and the brighter sample with spectra makes it possible to make discrimination among galaxy types and luminosities with much greater precision. The work presented by Tegmark and Zehavi at the Albuquerque AAS meeting, on the 3D power spectrum and the 3D two-point covariance function, respectively, shows error bars smaller than an identical analysis on the 2dF data set. The two functions, which formally are Fourier transforms of one another, but in practice are generated by wildly different techniques on somewhat different samples, agree impressively well when transformed.

The power spectrum analysis, which samples large scales rather better than the covariance one, shows convincingly that luminous galaxies are more strongly biased than faint ones when luminosity cuts are made. Since on large scales most of the weight comes perforce from distant and hence bright objects and on small scales the numbers are dominated by fainter objects, this results in a tilt of the power spectrum which is quite severe but which can be quantified and corrected. It looks currently like the bias variation results in a simple scale-independent amplitude shift with luminosity, but this must be checked carefully before the corrected results are given too much credence. This will be much more easily and accurately accomplished with the full sample at the end of the survey. These power spectrum analyses make use of linear theory checked with n-body simulations to remove redshift-space distortions, and the techniques work quite well. The analyses make use of expansions in Karhunen-Loeve eigenfunctions and are very efficient computationally.

On small scales (less than about 8 Mpc/h), the covariance analysis shows unambiguously that the covariance function varies both in shape and amplitude with different morphological types, and it is probable that no one type accurately traces the mass on such scales. This is astrophysically not surprising and was indeed an expected result from the SDSS's unique ability to do these analyses with morphological segregation, but will significantly complicate the interpretation of the results.

The use of statistics different and in some ways more powerful than the two-point analyses is also facilitated by the size and quality of the SDSS dataset; techniques to evaluate higher-order covariance functions in a computationally manageable way have been developed by Szapudi, and the topological techniques developed by Gott have been applied in a preliminary way by Hoyle and others, though the survey footprint at this time is not very satisfactory for these analyses. There is little enough experience with these tools at this time that it is difficult to know how powerful they will be, but they should at various scales put powerful constraints on the gaussianity of the underlying primordial density field, bias of the distribution of luminous matter in the universe relative to the dark matter, and provide direct tests of the notion that the structure in the universe arose from gravitational instability.

Essentially all of the investigations defined by the Key Projects list have been begun, and many of them are essentially as far along as the present quantity of data allow. The present results are in broad agreement with the currently fashionable 'concordance' lambda-dominated k=0 cosmological model, with Omega_m about 0.3, H ~ 70 km/sec/Mpc, and a baryon fraction in the vicinity of 0.2, as presented by Pope at the AAS meeting. As the quantity of data increases and the footprint on the sky becomes more nearly contiguous the error bars on essentially all quantities will decrease and the window functions for the power spectrum will become much better behaved. We (and 2dF) cannot yet meaningfully comment on the expected baryon wiggles in the power spectrum, the existence of which are a powerful clue to the general correctness of our cosmological picture, but we do now understand what the error bars will be at the end of the survey; if we obtain the full originally defined area of 10,000 square degrees we should easily resolve these features, and may be able to if we are forced to stop after five years with a roughly 7000 square degree footprint. Work on more complex statistics (particularly the characterization of voids and the topology work) will be vastly facilitated by a contiguous survey area.

6.2 Clusters of Galaxies

There has been a very large amount of work done on clusters, primarily at this point on the creation of cluster catalogs, the first item in the Key Projects list. The difficulty of characterizing clusters, a problem present essentially since the infancy of the subject, has made it difficult to proceed; different aggregates at similar scales and over densities are detected in quite different ways by different techniques and show very large dispersions in properties. Progress is being made on this problem, but slowly. Some questions of considerable astrophysical/cosmological importance, such as the mass function of clusters, have been addressed (by Bahcall et al), using both techniques anticipated at the creation of the Key Project list and new ones (lensing) not properly anticipated at the time. Further significant progress probably awaits a satisfactory resolution of the catalog/characterization issue, which is proving very difficult except for the very richest aggregates; these, however, because they represent extreme density enhancements, promise to be very powerful levers on the cosmological parameters or on the statistics of the density field depending on which direction one approaches the problem, and there are solid results on these questions already, as presented at the AAS by Sheldon et al and by Annis et al.

6.3 Quasars

Work has been going on for a long time on the two quasar key project areas, though papers on the luminosity function over the full redshift range and on the clustering will not appear for some months yet (Fan et al and Vanden Berk et al). This is an area, however, in which the most important and exciting science, which is certainly the work on very high-redshift objects, was not really anticipated. We had no way of knowing when the survey was conceived that we would be so successful at finding these objects and be able to develop effective techniques to overcome the substantial obstacles in the way of the search; these same techniques, due largely to Fan and Lupton, have produced as well the methane brown dwarf sample, as we discuss below.

We also, of course, had no way of knowing that very near our redshift limit of about 6 was the epoch of reionization, which has been one of the most exciting results to emerge from the SDSS. We have so far three objects, which effectively probe this epoch, with redshifts from 6.2 to 6.4 (the additional two objects with $z>6$, at just above 6, are not sufficiently distant to show the effect unambiguously). The agreement among the few lines of sight investigated so far is encouraging, but the number is still very small; the discouraging thing is that this sample is from a thorough search of the roughly 3000 square degrees currently in the can, and will still not be very large when we are done, though will be substantially larger if we cover the full 10,000 square degrees than if we stop after five years. It is important to note that for searches of rare objects like these very high-redshift quasars the power of a survey is *LINEAR* in the area covered, not the square root, as a naive interpretation of the statistical significance of large samples would suggest. We do know from the luminosity function work at high redshifts that the luminosity function is much shallower at these epochs than later, so going fainter will not yield large numbers of objects, and a wide-angle survey like the SDSS is the most efficient way to find them. It is also cogent that the objects we find are faint enough that they are already quite difficult to study with telescopes like Keck and the VLT, and fainter objects would be much more difficult yet at the signal-to-noise ratios required to study reionization.

A related area has touched by the work of Bernardi et. al. on the effect of the reionization of helium, as revealed by irregularities in the dependence of the average hydrogen absorption line density with redshift. The reionization itself cannot be observed in the optical (or even in the ultraviolet without very large telescopes) but the effect on the thermal history of the intergalactic medium is fairly profound and can be traced, as this work demonstrates.

An area of somewhat unexpected richness has been the large and incredibly varied sample of broad-absorption-line (BAL) quasars, studied by Hall et al, which are among the spectroscopically most bizarre objects ever studied in astrophysics. Their spectra are dominated, as their name suggests, by broad absorption features of many elements in many ionization states arising in fast outflows from the central engines in these objects. The spectra are at the present state of understanding more than a bit bewildering, but there is considerable promise that they will lead to a much better understanding of the conditions in these enigmatic objects.

6.4 Galaxies

Quite different from the study of the distribution of galaxies in space (but inextricably connected with it) is the study of the properties of galaxies themselves, to which the SDSS has made and continues to make fundamental contributions. For all of this work the size of the data set, control of selection, and quality of the SDSS data are paramount. Number counts in the imaging data and the related work on luminosity functions in the spectroscopic sample have been investigated by Yasuda et al and Blanton et al, respectively. The luminosity function work for the first time clarified the many rather loose ends which had developed over the years concerning the dependence of the LF on type, color, etc. At about the same time, the work of Strateva et al showed a clear dichotomy in the properties of galaxies as a function of u-r color, in which the distribution of galaxies at any magnitude is clearly bimodal and separated naturally into 'late' star-forming systems and 'early' almost passive systems. This work was done with still small early samples and has been vastly extended as the number of galaxies in the spectroscopic sample has increased.

Two more recent studies, still in final stages of preparation, are the work of Hogg, Blanton, et al on distribution functions of an important set of observed galaxy properties and the work in Tremonti's thesis on metal abundance. The former uses the complete main galaxy sample to construct distribution functions normalized to unit comoving volume and to the median redshift of the survey for color, luminosity, central concentration, and surface brightness, and the distribution of these quantities with local galaxy density. The results can be shown either in number, or, perhaps more informatively, as luminosity-weighted distributions. From these density plots and the related conditional probability distributions, which can be constructed from them, one can investigate the distributions and dependencies of almost any of the gross properties of galaxies on any other and on environment. The accurate photometry and well-controlled selection in the SDSS makes constructing believable functions of this sort possible, and the very large sample size is necessary to allow accurate determination of these multidimensional distribution functions.

Another dimension has been added to these studies in the work of Tremonti, Heckmann, Kauffman et al. Many galaxies have emission lines strong enough to apply the methods developed over the years to determine the chemical abundances of the gas in these systems. The result of this work underscores the dichotomy found in Strateva et al's investigation. The distribution of galaxies is grossly characterized by the division into two populations: One of massive, centrally concentrated, high surface brightness, red galaxies with high and relatively uniform metal abundance and moderate to low star formation rates and a more numerous population of less massive galaxies of lower luminosity and surface brightness, bluer colors, higher specific star-formation rates, and lower metal abundance which is highly variable from object to object but very strongly correlated with the stellar mass in the system. The metallicity-mass relationship is very tightly constrained and shows the strong and tight correlation with mass up to stellar masses of about 3×10^{10} solar masses and saturation at about solar abundance for higher masses. These findings are in qualitative agreement with the predictions of Larson twenty years ago of supernova-driven mass loss and have profound implications for current pictures of galaxy formation. In particular, one expects the baryon to dark matter ratio to drop precipitously with decreasing luminosity once one is discussing systems on this rising branch and thus the overall mass-to-light ratio to rise similarly. Low-luminosity systems certainly cannot be expected to trace halo mass, and the relationship between the luminosity function and mass function is complex; certainly there will be far fewer low-luminosity systems than one would naively expect assuming constant baryon-to-dark-matter ratios. Since the observations were already at variance with the CDM predictions in this regard these results can be regarded as promising steps toward solving this problem, but the observations are by now way ahead of theory.

What might we expect in these areas by the end of the survey? The results obtained to date are already far, far better than any past work has produced. The missing link which will be pursued as the data volume increases and one understands the distributions better is the connection of all of this with large-scale structure---that is, the spatial distribution of these properties. There is some considerable hope that the chemical studies will allow us to predict the halo mass associated with a galaxy at least approximately and give us a much better handle on the mass power spectrum and covariance function, and the behavior in space of these properties will strongly constrain notions of galaxy formation. It is difficult at this juncture to predict how this will go, but it is clear that even with the full million galaxy redshift sample we will be hard-pressed to study these distributions as fully as one would like. Weak lensing, as we will see below, offers a completely independent halo mass determining tool, and the combination of these techniques promises a powerful and unique tool to study the joint distribution of baryonic and dark matter in the universe connected with galaxies.

Thus the science we have done in the `core programs' has been fundamental and strongly guided by our original plans and the framing of the Key Projects. It would have been very disappointing, we believe, if a survey of this size and complexity had not in addition yielded many other important results. We have, indeed, not been disappointed. A short summary of results of these other investigations follows.

6.5 SDSS Science Beyond the Key Projects

The SDSS was designed quite deliberately to have a well-defined science agenda in order to narrow the requirements on the hardware and the software. This agenda, large-scale structure and statistical properties of galaxies and quasars, is captured in the Key Projects that we have discussed above. There are nevertheless a wealth of other scientific programs that can and have been addressed with the SDSS data. Some of these were described in the Project Book, but a number of others became apparent only after we had data to evaluate. In this section, we describe a number of important scientific directions the SDSS Collaboration has been pursuing that are outside of the Key Projects rubric.

6.5.1. Asteroids

The motion characteristic of a main-belt asteroid can be detected during the few minutes it takes to scan any particular point in the sky, and the amount of this motion is a measure of its distance from us. Accurate colors in the SDSS filters can be derived, which provide some information on the surface composition. Since the ecliptic is included within the SDSS footprint, the result is the detection of unprecedented large numbers of asteroids: at the end of the survey we will have discovered 100,000 asteroids, all with multi-color photometry. This work is being pursued by workers at several institutions, including Ivezic, Lupton, and Quinn. At present, over 10,000 SDSS detections have been matched with earlier discoveries; this is the sample for which we have accurate orbital information. SDSS has demonstrated that families of asteroids (defined by the orbital parameters of semi-major axis, inclination, and eccentricity), presumed to have originated from the same parent body, show similar colors. While this result confirms earlier work, the SDSS data have such dramatic statistical weight that one can now invert the logic: clumpings in color space can confirm the reality of clumpings in orbital parameter space. The SDSS data give us a new way to explore evolutionary processes in the Solar System; it now seems that most asteroids can be associated with a family.

One of the most dramatic results has been the demonstration that the size distribution of small main-belt objects is much less steep than thought previously. The implications of this for the size distribution of potentially dangerous near-earth objects is highly suggestive but still somewhat controversial. Recent improvements in the achieved astrometric precision demonstrate that techniques similar to those used to study the main belt objects can be employed to detect and study bodies out to at least the orbit of Neptune.

6.5.2. Structure in the Milky Way Halo

It has been commonplace to characterize the distribution of stars in the halo of the Milky Way as a smooth spheroid, using a small number of parameters to describe the geometry. When we look at other isolated galaxies, their spheroids usually look smooth (although some elliptical galaxies show shell structures). Since individual stellar orbits generally change very slowly, the distribution of orbits can tell us about the formation of galaxy halos.

The SDSS detects as many stars as it does galaxies; we will have photometry for over 100 million stars by the end of the survey. The colors in the SDSS filter system allow stars to be classified, and in particular to be assigned a distance based on the derived luminosity. In this way, the SDSS is building up a 3-D map of the distribution of stars in the Milky Way (mostly in the halo since we are not scanning low latitudes) that has much better precision (because of the quality of the photometry and the multiple bands) and much better statistical power than possible before. Instead of a smooth distribution of stars, we are finding that there are structures (clumpings) over a wide range of angular scales. The problems are being pursued with a variety of complementary techniques by Newberg, Yanny, Rockosi, Odenkirchen, and Grebel. On the scale of 10 degrees, we have detected tails--- long plumes of stars---associated with the tidal disruption of the globular cluster Palomar 5 by the gravitational potential of the Milky Way. The detailed character of the tails provides strong constraints on dynamical models: we now have a firm notion of the orbit of this cluster, and more significantly, we can place limits on the tidal field and thus on the amount and distribution of dark matter. In another series of studies on even larger angular scales, the tidal stream associated with the Sagittarius dwarf galaxy has been studied in much greater detail than before, and over much larger regions of the sky. Still other large-scale lumps in stellar density have been identified in the maps, which are presumably related to dwarf galaxies or globular clusters that have been completely disrupted. These discoveries lead to a completely different picture for the origin of the halo from the classical one: probably much of the mass in the halo was assembled from the accretion of small galaxies, with persistent inhomogeneities and other dynamical signatures. Further work will concentrate on obtaining radial velocities for stars in these structures, and in extending the maps to cover more of the Galaxy. An important parameter that is related to the distribution of dark matter is the degree of flattening of the halo. With further measurements, we expect to be able to measure this parameter reliably.

6.5.3 Brown Dwarf Stars

An object born with a mass less than 0.08 solar mass cannot stabilize itself by the ignition of fusion reactions in its core. Its thermal energy leaks away and the object fades on a relatively short time scale. Such objects, collectively called brown dwarfs, have sizes similar to that of Jupiter and at the low-mass end masses of only a few Jupiter masses. While these objects probably do not account for much of the mass in the solar neighborhood, it is still very important to determine the space density and other properties of brown dwarfs because we want to explore the continuity of formation processes from planets to stars. Because of the low luminosity of brown dwarfs and their relatively fast cooling times, to search for them it is essential to cover a large solid angle, even though their space density is relatively high. SDSS uniquely samples large angles to faint limits, and has had great success in the discovery of new brown dwarfs, both the relatively warm L dwarfs and the very cool (<1000K effective temperatures) T or methane dwarfs, which cannot be found from infrared colors alone. The combination of our optical colors with the near-infrared colors of the 2MASS survey has found even more brown dwarfs. Depending on the temperature and the chemical abundances, the spectra of these objects look quite bizarre (strong molecular absorption blocks out large ranges in wavelength), progressively becoming more like the spectra of giant planets as the temperature decreases; in the T spectral range the methane and water absorption is so strong that the infrared colors do not reflect the low effective temperatures at all, and detecting the virtual absence of flux in the spectral range below a micron is necessary for discovery. We detect the coolest of these objects only in our z band. They have thus been largely and perhaps ironically a byproduct of our search for the very highest redshift quasars, so the same search has yielded the most distant and brightest objects in the universe and the feeblest, coolest stars. Again Fan has led the search effort on the extreme objects, with work on these and less extreme objects which turn up with the normal survey sampling procedures by Knapp, Strauss, Golimowski, Hawley, and others. The ensemble of different types of spectra have helped fill in gaps between various classes, leading to a better understanding of the evolutionary cooling. SDSS found the first instance of a methane dwarf not in orbit around another star; once we have adequate statistics, constraints can be placed on formation scenarios. A complete sample has recently been constructed from our data by Collinge, Knapp, et al, and from this we expect to discover of order 100 T stars, which are bright enough to study adequately. Whether this sample is large enough to characterize them well is dubious, since they cover a fair range in masses and age, so this is another subject in which we gain essentially linearly with covered area.

6.5.4. A Zoo of other Weird Stars

We fully expected that there would be serendipitous discoveries with the SDSS, and by far the largest number of them have been a very large number of very unusual stars. These include very cool white dwarfs which are very blue because of the strange behavior of molecular opacity in their atmospheres, very cool very metal-poor main sequence stars, which have spectra entirely different from their more metal-rich cousins in the disk, dwarf carbon stars much hotter than have been seen before, a large variety of white dwarfs with very large magnetic fields, an M dwarf with some kind of compact magnetic companion with what are almost certainly cyclotron lines in the blue region of the spectrum. Many of these objects have had a few examples known previously, but no real information about densities, populations, etc.

6.5.5. Weak Lensing

The gravitational bending of light from a more distant object (the source) by an object in the foreground (the lens) can magnify and distort the image of the source. In this way, it is possible to obtain valuable information on the amount of mass along the line of sight, information that is very hard to obtain in any other way. Since the distortions are tiny, and the undistorted shape of the source is unknown, the technique is inherently statistical. Large numbers of galaxies must be measured down to a high surface density on the sky, and any distortions introduced by the atmosphere or telescope optics must be corrected with high precision. Once again, the SDSS data provide an exceptional new capability because of the volume and precision of the data. This effort has been led by McKay, Fischer, Seljak, Sheldon, and others. The lensing signal due to the massive halos of foreground galaxies has been detected to hundreds of kiloparsecs radial distance. Knowing both the mass and the luminosity, we can investigate how these quantities are correlated on the scale of galaxy halos. This, in turn, provides constraints on pictures for galaxy formation, namely how the baryons that fell into the dark matter potential wells lit up in the form of stars. The SDSS sample is so large that the lens galaxies can be subdivided into categories with different star-formation histories, different internal structures, and different environments. It appears that the scaling of light with mass is remarkably similar in the reddest bands, even across a wide range of galaxy types among relatively luminous galaxies. Eventually this work will push down to lower luminosities and can be combined with the work on metal abundances described above to paint an accurate picture of how baryonic matter is associated with dark matter halos.

7. Outreach and Communication

7.1 Graduate and Undergraduate Education

In the short time since the project delivered the first SDSS commissioning data to the collaboration, ten students have completed their PhD and begun careers in the disciplines that the SDSS touches. During the interval 1 January – 30 September 2002, Drs. M. Weinstein, E. Sheldon, and R. Scranton defended their dissertations, and the last two individuals moved on to SDSS institutions (The University of Chicago and the University of Pittsburgh, respectively). The current and future SDSS graduates are in an excellent position to emerge as leaders of the next generation of surveys.

As of 30 September 2002, 27 graduate students were working on dissertation topics based on SDSS data at 11 institutions (counting the Max Planck institutions as one). The list of these topics is presented in Appendix C; nine of these were posted in the interval 1 January – 30 September 2002. In the context of outreach, it is noteworthy that six of the 27 students are working with External Participants, scientists who are not on the staff of participating institutions.

The large quantity of superb, reduced SDSS data allow students, including undergraduates, to dive into real research with little overhead. The total number of undergraduates who have participated in a significant way in SDSS activities in the interval 1 January – 30 September 2002 is at least 33, representing three participating institutions and eight other institutions with External Participants. Of these students, nine conducted senior theses based on SDSS data. Eleven papers at AAS meetings featured undergraduate authors, and seven papers are in press in refereed journals.

7.2 Collaboration Meetings

More than two hundred people in the collaboration are actively working with the Archive. To share information and develop research teams, collaboration meetings are held twice a year, rotating among the participating institutions. Typically about sixty people participate in each meeting. The Collaboration Council selects the meeting location and serves as the scientific organizing committee. In 2002, the spring meeting (March 21 – 23) was held in Heidelberg. The second meeting in 2002 was at Princeton July 23-27. Both meetings included Working Group sub-meetings and plenary presentations of current research, as well as technical discussions, e.g. photometric calibration and testing for Data Release 1. The spring 2003 meeting will be held in Flagstaff. The participating institutions and ARC make a special effort to provide support for graduate students and younger members of the collaboration to attend these meetings and make presentations of their work.

In addition to the Collaboration meetings, the Working Groups have been encouraged to organize workshops on an ad hoc basis. In January there was a workshop at the University of Arizona on the subject of galaxy spectral synthesis, including members of the Large-Scale Structure Working Group, the Galaxies Working Group, and the Clusters Working Group. In September, the Stars Working Group held a workshop on white dwarfs in Flagstaff.

7.3 American Astronomical Society Meeting

The SDSS was represented at the summer AAS meeting in Albuquerque where we organized a Special Session on early results in the field of Large-Scale Structure from the SDSS. The summer meeting also featured a press conference (and associated press release) on the 10-degree tidal tail of Pal 5. The use of the SDSS databases and the mining tools (including the SkyServer) were demonstrated at the SDSS exhibit booth. About 20 media professionals (mostly science writers) toured Apache Point Observatory at the conclusion of the meeting.

7.4 Outreach at Apache Point Observatory

Guided tours of the SDSS facilities were afforded to 20 groups, on request, this past year. The groups consisted of school children, engineers, amateur astronomers, and personnel from Holloman Air Force Base and White Sands Missile Range. Also, several groups of students from various ARC institutions visited the site and were given detailed tours of the SDSS.

A 5-minute informational video was produced for use in the Sunspot Visitor and Astronomy Center, as well as distribution to SDSS and ARC member institutions and the public. The video focuses on each telescope located at Apache Point and discusses the science accomplished. Approximately 100 copies have been distributed to date.

NMSU is finishing a 30-minute documentary on SDSS for PBS distribution next spring. Filming was completed last winter.

7.5 Public Information Officer

Mr. Gary Ruderman, a free-lance writer based in Chicago, was hired as the SDSS Public Information Officer in May for part-time work to provide a central point-of-contact for various aspects of SDSS outreach. He has visited several of the SDSS participating institutions, as well as APO, to meet with the respective institutional public information officers and to foster a team identity. He will coordinate future press releases, help with activities at AAS meetings, and help update the www.sdss.org site.

7.6 Web Sites and Outreach

The SDSS web site http://www.sdss.org provides the status of the survey, links to published papers, policies for governance and publication, and access to the Archive. It also provides the most recent press releases on research results obtained by the SDSS Collaboration.

An important outreach element of the SDSS is the SkyServer http://skyserver.fnal.gov/en/, which provides astronomy education materials to students and their teachers in the context of access to SDSS data. SkyServer has more than 15 lesson plans for intermediate school, high school, and undergraduate students. In each lesson, students look at and analyze SDSS data to learn interactively. Each lesson includes a teacher's page, with sample student responses, teaching advice, and correlations to national education standards. In 2001, the number of page views for the SkyServer totaled 780,849, for a per-month average of 111,549 (June through December). Our 2002 page views were 1,473,313, for a per-month average of 163,701 (January through September), indicating a 45% increase in page views from 2001 to 2002. In the interval from January through September 2002, fifty-four teachers registered with the site.

Developers of the SkyServer attended the following meetings and made presentations: 1) March 1, Grid Physics Network (GriPhyN) outreach meeting, Brownsville. GriPhyN is setting up an outreach program and want to use SkyServer as a model. 2) June 12-14, NSAS Office of Space Science and Outreach meeting, Chicago.3) June 26-30, Space Science Workshop at the Wright Center for Science Education, Tufts University, Boston. The workshop focused on cosmology. Jordan Raddick led a workshop using SDSS data served by the SkyServer for about 30 teachers. 4) July 12, National Virtual Observatory outreach meeting, Baltimore. This meeting was to develop a "wish list" for NVO outreach - what do outreach professionals want NVO to be able to do? 5) August 3-7, American Association of Physics Teachers, Boise. Rob Sparks and Jordan Raddick led a half-day workshop on SkyServer for 30 high school, community college, and college teachers from around the country.

8. Financial Performance: 2002 Budgets and Costs

The operating budget, which the Advisory Council approved in November 2001, for the year 2002 consisted of $2,291K of in-kind contributions from Fermilab, US Naval Observatory (USNO), Los Alamos National Laboratory (LANL), and the Japan Participation Group (JPG); and $3,425K for ARC funded expenses. The sources of funds for the 2002 budget are shown in Table 8.1.

Table 8.1. Sources of Funds for the 2002 Budget ($K)

Sources of Funding	Cash	Actual In-Kind	Total
A. P. Sloan Foundation	2,000		2,000
Japan Participation Group	205	5	210
National Science Foundation	955		955
Prior year funds	265		265
Fermilab		1,580	1,580
Los Alamos National Laboratory		257	257
United States Naval Observatory		117	117
Total	3,425	1,959	5,384

The initial ARC-funded cash budget of $3,425K included $200K for management reserve. A review of our cost performance through September 2002 predicts that expenses for 2002 will be $3,256K. Since this is below the approved budget, we reduced the management reserve to $69K and have allocated $100K of the unused funds from the 2002 budget to the management reserve in calendar year 2003. Table 8.2 shows the forecast cost performance by project area for ARC-funded cash expenses in 2002. The forecast is based on actual expenses through September, and estimated expenses for the period October through December.

Table 8.2. Summary of ARC-Funded Cash Expenses for 2002, by Project Area ($K)

Category	Budget (Nov 2001)	Predicted Actual Expenses[1]
Survey Management	249	284
Collaboration Affairs	16	8
Observing Systems	870	842
Data Processing and Distribution	641	695
Observatory Support	1,360	1,360
ARC – Corporate Expenses	88	67
Sub-total	3,225	3,256
Management Reserve	200	69
Total	3,425	3,325

1. *Includes actual expenses through Q3 and predicted expenses for Q4.*

The cost increase in Survey Management is due to teleconference charges and support costs for the SDSS Public Information Officer. The cost increase in Data Processing and Distribution is due to the purchase of computing hardware to support Data Release 1. The costs for these needs were not included in the 2002 budget, when it was prepared in November 2001. The specific requirements for computing hardware were not known until May 2002. We allocated funds from the existing management reserve to cover these costs. After the books for 2002 are closed, we will place any remaining unspent management reserve funds in the management reserve for 2004 and beyond.

Table 8.3 compares the budgeted and actual in-kind contributions in 2002 by institution and Table 8.4 shows the distribution of in-kind contributions by project area. The estimated value of the in-kind contributions in 2002 is $1,959K. In-kind contributions were $332K below the 2002 budget for the following reasons: The level of in-kind support provided by Fermilab to support the observing systems at APO was less than forecast, in part because the amount of actual support required for existing systems was less than we estimated, and in part because the high priority that Fermilab placed on its collider program reduced the availability of some people from the Beams Division. The level of in-kind support provided by the JPG was less than forecast in the 2002 budget, because the actual effort required to support the imaging camera was less than anticipated. These reductions were partially offset by a slight increase in the in-kind contributions made by LANL. LANL provided engineering support to expand and maintain the capability of the telescope performance monitor, and astronomer support for the testing effort associated with DR1 preparations. The USNO group provided the project with the agreed-upon level of effort.

Table 8.3. Budget and Costs for the 2002 In-kind Contributions, by Institution ($K)

Institution	Budget (Nov 2001)	Actual In-kind Contribution[1]
Fermilab	1,903	1,580
Los Alamos National Laboratory	231	257
United States Naval Observatory	117	117
Japan Participation Group	40	5
Total	2,291	1,959

1. Includes actual expenses through Q3 and predicted expenses for Q4.

Table 8.4. Summary of In-kind Contributions, by Project Area ($K)

Category	Budget (Nov 2001)	Actual In-kind Contribution[1]
Survey Management	196	213
Observing Systems	969	715
Data Processing and Distribution	1,126	1,031
Total	2,291	1,959

1. Includes actual expenses through Q3 and predicted expenses for Q4.

The costs reported for 2002 are preliminary, since this report was prepared prior to the end of the calendar year. The preliminary costs are based on actual expenses through Q3, and a revised forecast of Q4 expenses made by the participating institutions. As actual costs incurred by some institutions through December will not be reported until the end of the first quarter of 2003, we will amend this report with the final cost accounting at that time. Details on the use of funds obtained from the Sloan Foundation and the National Science Foundation are provided in Appendix A.

9. Financial Planning

9.1. 2003 Budget

On November 25, 2002, a budget of $5,200K for the year 2003 will be presented to the ARC Board of Governors for approval. The proposed 2003 budget is fully funded. It consists of $3,400K in cash provided by ARC, including a management reserve of $213K, and in-kind support from the MOU Partners with an estimated value of $1,800K. The sources of funds for the 2003 budget are shown in Table 9.1.

Table 9.1. Sources of Funds for the 2003 Budget ($K)

Sources of Funding	Cash	In-Kind	Total
A. P. Sloan Foundation	2,000		2,000
National Science Foundation	885		885
Japan Participation Group	175		175
Prior year funds	340		340
Fermilab		1,500	1,500
Los Alamos National Laboratory		155	155
United States Naval Observatory		125	125
Japan Participation Group		20	20
Total	3,400	1,800	5,200

The funds from the A.P. Sloan Foundation are from a commitment to award ARC $10,000K for the observation phase of the Five-Year Baseline Survey. The initial award was made in December 1999 and to date the A.P. Sloan Foundation has awarded ARC $7,000K for the Five-Year Baseline Survey. We anticipate that they will provide ARC with $2,000K at the end of the calendar year. The award of the remaining $1,000K for 2004 will be subject to making satisfactory progress on the survey. ARC anticipates receiving $885K from the National Science Foundation in 2003, as shown in Table 9.1. This will be the third allocation from a multiyear grant of $4,000K that ARC received in August of 2001. The amount of NSF funds that ARC will receive in 2003 and beyond is subject to fulfilling the terms of the grant. The funds from the Japan Participation Group represent their commitment to purchase $175K in supplies and they will be primarily used to finance finished plug plates.

With regard to in-kind contributions, Fermilab will continue to provide support for Observing Systems, data processing and distribution, and survey management. Los Alamos will provide support for the maintenance of the observers' software and testing the output of the data processing pipelines. The U.S. Naval Observatory will provide support for the maintenance of the astrometric pipeline and the operational database, and the Japan Participating Group will provide technical support for the imaging camera.

Table 9.2 shows the allocation of the 2003 cash funds by project area and compares the 2003 budget to the actual 2002 cash expenditures in each category.

Table 9.2. Allocation of 2003 Cash Funds, by Project Area ($K)

Category	2002 Actual Expenses*	Proposed 2003 Cash Budget
Survey Management	284	227
Collaboration Affairs	8	34
Observing Systems	842	769
Data Processing and Distribution	695	533
Observatory Support	1,360	1,447
ARC – Corporate Expenses	67	177
Sub-total	3,256	3,187
Undistributed Contingency	69	213
Total	3,325	3,400

*Includes actual expenses through Q3 and predicted expenses for Q4.

Table 9.3 shows the distribution of anticipated in-kind contributions by project area, along with a comparison of the estimated value of actual 2002 in-kind contributions.

Table 9.3. Distribution of 2003 In-kind Contributions, by Project Area ($K)

Category	2002 Actual Expenses*	Proposed 2003 Budget
Survey Management	213	216
Observing Systems	716	562
Data Processing and Distribution	1,031	1,022
Total	1,959	1,800

*Includes actual expenses through Q3 and predicted expenses for Q4.

9.2. Financial Planning: Funding Requirements

The estimate of $28 million for the five-year observing phase of the Survey has been revised to include costs for a modest project closeout plan. The closeout plan provides support for a skeleton crew to complete an orderly shutdown of SDSS operations over a three-month period beginning in July 2005. Shutdown activities include decommissioning systems to leave them in a safe state, completing final documentation, writing final summary reports, and paying final invoices. They also include processing any unprocessed data and transferring the SDSS Archive to a long-term steward. The initial draft of the closeout plan was presented to the Advisory Council in May 2002 and is currently undergoing refinement. The estimated cost of the closeout is $341K, of which $223K is cash and the remainder is an in-kind contribution from Fermilab. The estimated cost of operations for 2003 through the end of the five-year survey, including the closeout, is $13,168K.

We added two new partners in 2002. In the first quarter of 2002, we reached agreement on an MOU with Los Alamos that gives a limited number of Los Alamos scientists access to the SDSS data for specific projects prior to its release to the astronomical community in exchange for the valuable in-kind contribution that Los Alamos has made over the last several years and their commitment to continue to provide this in-kind support at its current level. In the second quarter, the University of Pittsburgh joined the SDSS as an Affiliate Member by making a one-time cash payment of $550,000 to the project. They have also agreed to make an in-kind contribution in support of science testing. With these commitments and the commitments from the existing sponsors, ARC can complete the five-year survey.

Since the five-year survey is drawing to a close, the Advisory Council has begun an evaluation of the scientific capabilities of the SDSS systems in the period after June 2005, in light of competing capabilities of other observatories. At the October meeting, several Council members presented proposals for possible uses for the 2.5m telescope and instruments in that period at the October meeting. Several members presented the scientific case for continuing operations in the current mode to fill in the gap in the imaging survey of the Northern Galactic Cap that is likely to exist at the end of June 2005. Several members advocated a dedicated imaging survey of a small area of the sky in order to study the time dependence of objects such as supernovae and asteroids. On the basis of the proposals, it appears that the SDSS equipment will remain a unique and powerful asset during the period from 2005 through 2010. After hearing these presentations, the Council established a "Futures" committee and charged them with developing a scientific scope and operations cost estimates for several competitive future programs. The chair, Suzanne Hawley, will present the initial recommendations of the committee at the annual meeting of the Advisory Council.

10. Outlook

Our ability to acquire survey quality data when the atmospheric seeing meets survey requirements improved substantially during the past two years. The Observing Systems are performing well and are proving to be very reliable. We plan to improve on that performance in 2003. The SDSS Engineering group, with the support of the SDSS Observers, and the APO staff will place greater emphasis on the creation and implementation of a scheduled maintenance plan. They will also assemble a complete inventory of critical spares. During the past years, the Observers improved their operating procedures and they plan to continue this effort in the coming year, while placing greater emphasis on documenting those procedures. We believe that with careful attention to maintenance and operation procedures, we can sustain the excellent performance the Observing Systems to the end of the survey. The Observatory Staff has met the needs of the SDSS and the infrastructure improvements that they guided to completion in 2002, such as acquiring bigger and better outfitted office trailers and the new phone system will contribute to the efficiency of operations continue to keep morale on the mountain high.

The astronomers form the Participating institutions made a major investment in improving the scientific software and testing the outputs of the software. Much of that effort went into improving the performance of the major pipelines, Photo and Spectro, and it led to significant improvement of the quality of the processed data, particularly photometry and the astrometry. These improvements have already strengthened the value of the science that was derived from the data. As the pipelines became stable in 2002, it was possible to develop a strong science testing program and this effort will continue through 2003 and beyond. We plan to develop a formal quality assurance program for the processed data. The challenge will be to make it coherent over the geographically dispersed collaboration.

One of only two disappointments in 2002 was the fact that we did not complete the development of our proposed public data distribution system. We did make great headway on building a high quality distribution system and we are committed to completing it in 2003. Once we have built the quality with the processed data, databases, and the distribution tools, we will distribute the data to the astronomical community. We will support the data release through a help desk. Our experience with the EDR persuaded us that we should focus on putting the quality into the data set, rather than meeting a specific deadline. However, whatever we do must be done within the envelope of our limited financial resources.

The second disappointment that we sustained in 2002 was the slow rate at which we obtained data. We learned, as others have learned before us, that the weather is not only unpredictable, but it is unpredictably disappointing. In order for us to achieve the original goals that were laid out in the Principles of Operation (PoO) we will require at least a two-year extension to the five-year baseline survey. A three-year extension would provide a small contingency to compensate for weather. Since the achievement of the PoO goals will require an extension, the Advisory Council has created a Futures Committee that will examine the scientific opportunities that the SDSS Observing System could create after the five-year baseline survey is completed. The committee will assess the scientific merit of an extension as well as the scientific merit of new programs. It will report its findings to the Advisory Council in 2003.

A very important objective for 2003 will be to demonstrate the scientific power of the SDSS even more convincingly than we demonstrated in 2002. In 2003, the SDSS Collaboration of nearly 200 astronomers will be using the SDSS Archive for science. This will certainly produce some remarkable results, since the size and quality of the SDSS data now exceeds that of its predecessors.

APPENDIX A. Use of Funds from the A.P. Sloan Foundation and the National Science Foundation in 2002.

The funds from the A.P. Sloan Foundation, NSF, and other sources were expended in 2002 on the project areas shown in Table A1.

Table A1. Summary of Sloan, NSF, and Other Funded Expenditures on 2002 Costs (Accrual Basis, $K)

Category	A.P. Sloan Funds	NSF Funds	Other Funds	Total
Survey Management	235	49	0	284
Collaboration Affairs	5	0	0	8
Observing Systems	479	206	156	841
Data Processing and Distribution	510	184	0	694
Observatory Support (APO/NMSU)	699	516	145	1,360
ARC Corporate Expenses	67	0	0	67
Total Expenditures	2,000	955	301	3,256

The expenditures in Table A1 are for payments made by ARC prior to January 1, 2003. All of the NSF funds awarded to ARC in 2002 were fully committed by the end of calendar year 2002. Details of the payments from the A.P. Sloan Foundation, NSF, and other cash accounts between January 1, 2002 and December 31, 2002 are shown in Table A2. Each line lists the funds paid to a particular institution for the specific scope of work defined in the annual agreements between each institution and ARC. Each agreement has a specific SSP number and provides an initial budget for that SSP. During the year, the Project Manager and the Business Manager track the costs incurred by each SSP. They review all new commitments with the appropriate budget officer from each institution. All commitments in excess of $3,000 require the approval of the Director and the ARC Business Manager. Any changes in the personnel supported by ARC require a revision to the Agreement.

Some of the items deserve a few comments. The costs incurred by Fermilab for SSP48 are for travel in support of survey management responsibilities. Costs incurred by Fermilab on SSP42 include the salary of an electronics technician in permanent residence near APO, and the procurement of parts and materials, including spares, for the 2.5-m telescope and for all other hardware systems in Observing Systems except the DAQ system. SSP42 also provides funds for travel expenses for the Fermilab employees stationed at APO. Costs incurred by Fermilab on SSP40 include the salary costs for individual who maintains the SDSS website and is supporting the DR1 and computer hardware required to support Data Release 1. Costs incurred by Fermilab on SSP61 include the salary costs for an individual responsible for maintaining the observers programs at APO and for assisting in time tracking and efficiency performance analysis, and the travel costs for this individual to travel between Fermilab and APO. Reimbursement of all Fermilab costs is made from the A.P. Sloan Foundation account, as shown in Table A.2.

Table A2. Details of Sloan, NSF, and Other Funded Expenditures in CY2002
(Accrual Basis, $K)

SSP No.	Description	A.P. Sloan Funds	NSF Funds	Other Funds	Total
Survey Management					
SSP21	ARC Secretary Treasurer	18	0	0	18
SSP34	ARC Business Manager	56	0	0	56
SSP46	PU Support for Project Management	69	49	0	118
SSP64	NYU Photometric Calibration	4	0	0	4
SSP48	FNAL Survey Management	32	0	0	32
SSP65	UC Project Spokesperson	5	0	0	5
SSP91A	ARC Support for Public Affairs	26	0	0	26
SSP91H	SDSS Public Information Officer	24	0	0	24
	Subtotal	235	49	0	284
Collaboration Affairs					
SSP91B	ARC Spokesperson Support	0	0	0	0
SSP91C	ARC Support for Collaboration Affairs	8	0	0	8
	Subtotal	8	0	0	8
Observing Systems					
SSP42	FNAL Telescope Engineering Support	124	0	0	124
SSP61	FNAL DA & Observers' Program Support	64	0	0	64
SSP31	UW Telescope Engineering	0	170	132	302
SSP32	PU Hardware Support	85	18	0	103
SSP36	JHU Spectrograph Hardware Support	30	0	0	30
SSP33	UC Engineering Support	16	0	0	16
SSP91D	ARC Observing Systems Support	160	18	24	203
	Subtotal	479	206	156	841
Data Processing and Distribution					
SSP40	FNAL Software and Data Processing Support	177	0	0	177
SSP38	PU Software Support	192	0	0	192
SSP39	UC Science Software Support	41	0	0	41
SSP62	JHU End to End Testing and Validation	33	0	0	33
SSP63	UW End to End Testing and Validation	54	0	0	54
SSP37	JHU Data Archive Development and Support	13	184	0	197
	Subtotal	510	184	0	694
Observatory Support					
SSP35	NMSU Site Support	699	516	145	1,360
	Subtotal	699	516	145	1,360
ARC Corporate Support					
SSP91E	ARC Corporate Support	61	0	0	61
SSP91F	ARC Additional Scientific Support	0	0	0	0
SSP91I	SDSS Observers Research	6	0	0	6
	Subtotal	67	0	0	67
Total Expenditures		2,000	955	301	3,256

Table A3. Details of Sloan, NSF, and Other Funded Corporate Expenditures in CY2002 (Accrual Basis, $)

Code	Description	A.P. Sloan	NSF	Other	Total
SSP91A	AAS MEETING BOOTH	1,600	0	0	1,600
SSP91A	PRINT SW OBS POSTERS	896	0	0	896
SSP91A	IDIT ZEHARI (AAS TRAVEL)	1,054	0	0	1,054
SSP91A	JORDAN RADDICK (AAS COPIES)	498	0	0	498
SSP91A	PERFORMANCE GRAPHICS	10,000	0	0	10,000
SSP91A	AAS MEETING BOOTH	2,420	0	0	2,420
SSP91A	SDSS DR1 BROCHURES	10,000	0	0	10,000
SSP91A	ARC Support for Public Affairs	26,468	0	0	26,468
SSP91C	SUZANNE HAWLEY, TRAV ADV	700	0	0	700
SSP91C	SUZANNE HAWLEY, TRAVEL BALANCE	477	0	0	477
SSP91C	SCOTT ANDERSON, TRAVEL TO GERMANY	1,417	0	0	1,417
SSP91C	REN POLY TECH, HEIDI NEWBERG TRAV	1,110	0	0	1,110
SSP91C	TRAVEL REIMB TO COCO, NEWBERG	1,110	0	0	1,110
SSP91C	SUZANNE HAWLEY (TRAVEL REIMB)	344	0	0	344
SSP91C	SUZANNE HAWLEY (TRAVEL REIMB)	690	0	0	690
SSP91C	WG TRAVEL	1,000	0	0	1,000
SSP91C	TECHNICAL PAPER	1,500	0	0	1,500
SSP91C	ARC Support for Collaboration Affairs	8,347	0	0	8,347
SSP91D	MALORY STORAGE (JAN/MAR)	444	0	0	444
SSP91D	MALORY STORAGE (APR/JUN)	444	0	0	444
SSP91D	AURA, ALMINIZATION DEPOSIT	0	5,500	0	5,500
SSP91D	RENTSCHULER ELECTRIC	3,260	0	0	3,260
SSP91D	PAUL SPENCER (METALURGIST)	1,995	0	0	1,995
SSP91D	GENTRY CONST. (LOUVERS)	6,473	0	0	6,473
SSP91D	UW ASTRONOMY 2002 V-MILL USE FEE	0	0	24,175	24,175
SSP91D	COYOTE CABLING	42,104	0	0	42,104
SSP91D	GENTRY CONST, TRAILER SETUP	65,399	0	0	65,399
SSP91D	MALORY STORAGE (JUL/SEP)	444	0	0	444
SSP91D	SEHI COMPUTER PRODUCT (TEL SYST)	2,991	0	0	2,991
SSP91D	RAYTHEON INFRARED	0	5,951	0	5,951
SSP91D	GE CAPITAL (TRAILER DEMOBILIZATION)	1,305	0	0	1,305
SSP91D	GE CAPITAL (TRAILER FLOOR REPAIR)	475	0	0	475
SSP91D	OPTICOMM (TEL SYST EQUIP)	453	0	0	453
SSP91D	WILLIAMS-SCOTSMAN (ENG TRAILER)	5,882	0	0	5,882
SSP91D	WILLIAMS-SCOTSMAN (OBS TRAILER)	5,160	0	0	5,160
SSP91D	W-S (AUG ENG TRAILER RENT)	774	0	0	774
SSP91D	W-S (AUG OBS TRAILER RENT)	506	0	0	506
SSP91D	TRAILER RELATED P.CASH EXPENSES	5,000	0	0	5,000
SSP91D	TRAILER RELATED P.CASH EXPENSES	2,000	0	0	2,000
SSP91D	MALORY STORAGE (OCT/DEC)	444	0	0	444
SSP91D	W-S (SEPT ENG TRAILER RENT)	813	0	0	813
SSP91D	W-S (SEPT OBS TRAILER RENT)	541	0	0	541
SSP91D	TRAILER RELATED P.CASH EXPENSES	3,500	0	0	3,500
SSP91D	TRAILER RELATED P.CASH EXPENSES	1,905	0	0	1,905
SSP91D	ISLAND COMPUTERS 2-DS10L	2,105	0	0	2,105
SSP91D	W-S (OCT OBS TRAILER RENT)	511	0	0	511
SSP91D	W-S (OCT ENG TRAILER RENT)	775	0	0	775
SSP91D	GENTRY CONST. (TRAILER LANDSCAPE)	2,113	0	0	2,113
SSP91D	AURA RE-ALMINIZATION	0	5,500	0	5,500
SSP91D	W-S (NOV ENG TRAILER RENT)	520	0	0	520
SSP91D	W-S (NOV OBS TRAILER RENT)	520	0	0	520
SSP91D	DIMM ELECTRICAL, RENTSCHLER	0	1,250	0	1,250
SSP91D	W-S (DEC ENG TRAILER RENT)	775	0	0	775
SSP91D	W-S (DEC OBS TRAILER RENT)	520	0	0	520
SSP91D	ARC Observing Systems Support	160,149	18,201	24,175	202,525

Table A3. Details of Sloan, NSF, and Other Funded Corporate Expenditures in CY2002 - Continued
(Accrual Basis, $)

Code	Description	A.P. Sloan	NSF	Other	Total
SSP91E	BANK CHARGE	25	0	0	25
SSP91E	COMM LIAB INSURANCE	864	0	0	864
SSP91E	PETTY CASH FOR JAN-02 EXPENSES	2,578	0	0	2,578
SSP91E	PETTY CASH FOR FEB-02 EXPENSES	3,547	0	0	3,547
SSP91E	TRAVEL ASSISTANCE FUND	10,000	0	0	10,000
SSP91E	NOV-01 EXTERNAL REV MTG ($1,468)	0	0	0	0
SSP91E	BANK FEE	13	0	0	13
SSP91E	PETTY CASH FOR MAR-02 EXPENSES	2,534	0	0	2,534
SSP91E	DIRECTORS LIAB INSURANCE	5,882	0	0	5,882
SSP91E	NOV-01 EXTERNAL REV MTG ($2,514)	0	0	0	0
SSP91E	WIRE TRANSFER FEE	13	0	0	13
SSP91E	PETTY CASH FOR APR-02 EXPENSES	1,950	0	0	1,950
SSP91E	BANK FEE	17	0	0	17
SSP91E	PETTY CASH FOR MAY-02 EXPENSES	4,696	0	0	4,696
SSP91E	CLARK NUBER (AUDIT)	5,581	0	0	5,581
SSP91E	HILTON	3,187	0	0	3,187
SSP91E	MIKE EVANS (REIMB SDSS DINNER)	351	0	0	351
SSP91E	PETTY CASH FOR JUN-02 EXPENSES	3,000	0	0	3,000
SSP91E	D. YORK (TRAVEL REIMB)	93	0	0	93
SSP91E	CHECK PRINT CHARGE	64	0	0	64
SSP91E	PETTY CASH FOR JUL-02 EXPENSES	650	0	0	650
SSP91E	BANK FEE	0	0	0	0
SSP91E	PETTY CASH FOR AUG-02 EXPENSES	0	0	0	0
SSP91E	PETTY CASH FOR SEP-02 EXPENSES	500	0	0	500
SSP91E	PETTY CASH FOR OCT-02 EXPENSES	3,000	0	0	3,000
SSP91E	HILTON 10/15/02	2,913	0	0	2,913
SSP91E	MIKE EVANS (REIMB SDSS DINNER)	295	0	0	295
SSP91E	PETTY CASH FOR NOV-02 EXPENSES	3,000	0	0	3,000
SSP91E	PETTY CASH FOR DEC-02 EXPENSES	3,000	0	0	3,000
SSP91E	BANKING FEES	30	0	0	30
SSP91E	Hilton 11/25/02 (incl dinner)	3,200	0	0	3,200
SSP91E	ARC Corporate Support	$60,983	$0	$0	$60,983
SSP91H	G. RUDERMAN (SDSS PIO)	3,274	0	0	3,274
SSP91H	G. RUDERMAN (SDSS PIO)	1,258	0	0	1,258
SSP91H	G. RUDERMAN (SDSS PIO)	4,931	0	0	4,931
SSP91H	G. RUDERMAN (SDSS PIO)	5,000	0	0	5,000
SSP91H	G. RUDERMAN (SDSS PIO)	5,000	0	0	5,000
SSP91H	G. RUDERMAN (SDSS PIO)	5,000	0	0	5,000
SSP91H	SDSS Public Information Officer	$24,463	$0	$0	$24,463
SSP91I	SDSS OBS RESEARCH TRAVEL	2,239	0	0	2,239
SSP91I	SDSS OBS RESEARCH TRAVEL	2,000	0	0	2,000
SSP91I	SDSS OBS RESEARCH TRAVEL	2,000	0	0	2,000
SSP91I	SDSS OBS RESEARCH TRAVEL	260	0	0	260
SSP91I	SDSS Observers Research	$6,499	$0	$0	$6,499
	Annual Total	$286,909	$18,201	$24,175	$329,285

APPENDIX B - PUBLICATIONS

Q1 2002

Stellar Masses and Star Formation Histories for 80,000 Galaxies from the Sloan Digital Sky Survey
MRNAS submitted – Guinevere Kauffmann, et al.

Composite Luminosity Functions of the Sloan Digital Sky Survey Cut & Enhance Galaxy Cluster Catalog
PASJ submitted – Tomo Goto, et al.

Exploratory Chandra Observations of the Three Highest Redshift Quasars
ApJL, 569, 5 (2002) – W.N. Brandt, et al.

Optical and Radio Properties of Extragalactic Sources Observed by the FIRST Survey and the SDSS
AJ submitted – Zeljko Ivezic, et al.

Characterization of M, L and T Dwarfs in the Sloan Digital Sky Survey
AJ accepted – Suzanne L. Hawley, et al.

Comparison of Asteroids Observed in the SDSS with a Catalog of Known Asteroids
AJ submitted – Mario Juric, et al.

Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample
AJ submitted – Michael Strauss, et al.

Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Quasar Sample
AJ accepted – Gordon Richards, et al.

Q2 2002

Two-Dimensional Topology of the Sloan Digital Sky Survey
ApJ submitted – Fiona Hoyle, et al.

The Cluster Mass Function from Early SDSS Data: Cosmological Implications
ApJ submitted – Neta A. Bahcall, et al.

The Redshift of the Lensing Galaxy in PMN J0134-0931
ApJL accepted – Pat Hall, et al.

SDSS 0924+0219: an Interesting "Three Component" Gravitationally Lensed Quasar
AJ submitted – Naohisa Inada, et al.

SDSS 1226$-$0006: A Gravitationally Lensed Quasar Candidate from the Sloan Digital Sky Survey
AJ submitted – Naohisa Inada, et al.

Estimating Fixed Frame Galaxy Magnitudes in the SDSS
AJ submitted – Michael R. Blanton, et al.

Astrometric Calibration of the Sloan Digital Sky Survey
AJ submitted – Jeffrey R. Pier

The Application of Photometric Redshifts to the SDSS Early Data Release
AJ submitted – Istvan Csabai, et al.

Kinematic study of the disrupting globular cluster Palomar 5 using VLT spectra
AJ accepted – M. Odenkirchen, et al

The Dependence of Star Formation History and Internal Structure on Stellar Mass for 80,000 Low Redshift Galaxies
MNRAS submitted – Guinevere Kauffmann, et al.

Galaxy Star-Formation as a function of Environment in the Early Data Release of the Sloan Digital Sky Survey
ApJ submitted – P. Gomez, et al

The Sloan Digital Sky Survey Moving Object Catalog
– Z. Ivezic, et al.

A feature at z ~ 3.3 in the evolution of the Ly-alpha optical depth
AJ submitted – M. Bernardi, et al.

The Luminosity Density of Red Galaxies
AJ, 124:646 (2002) – D. Hogg, et al

Faint High Latitude Carbon Stars Discovered by the SDSS: Methods and Initial Results
AJ accepted – B. Margon, et al.

Color Confirmation of Asteroid Families
Nature submitted – Zeljko Ivezic, et al.

Cluster detection from surface-brightness fluctuations in SDSS data
A&A 388, 732-740 (2002) – Matthias Bartelmann, et al.

Cosmological Information from Quasar-Galaxy Correlations induced by Weak Lensing
A&A 386, 784-795 (2002) – Brice Menard, et al.

Based on public SDSS data

Constraining the Redshft z-6 Quasar Luminosity Function Using Gravitational Lensing
ApJ submitted – Zoltan Haiman

Detection of He II reionizastion in the SDSS quasar sample
ApJL accepted – T. Theuns, et al.

A Constraint on Gravitational Lensing Magnification and Age of the Redshft z-6.28 Quasar SDSS 1030+0524
ApJL submitted – Zoltan Haiman

A Physical Model for the luminosity of High-Redshift Quasars
ApJ submitted – Stuart Wyithe and Abraham Loeb

Morphological Butcher-Oemler effect in the SDSS Cut & Enhance Galaxy Cluster Catalog
PSAJ submitted – Tomo Goto, et al.

SDSS Survey for Resolved Milky Way Satellite Galaxies II: High Velocity Clouds in the EDR
AJ submitted – Beth Willman, et al.

The Sloan Digital Sky Survey
Contemporary Physics accepted - Jon Loveday, et al.

Broad Emission Line Shifts in Quasars: An Orientation Measure for Radio-Quiet Quasars?
AJ in press – Gordon T. Richards, et al.

Large Scale Structure in the SDSS Galaxy Survey
MNRAS submitted – Andrei Doroshkevich, et al.

Revision of the selection function of the Optical Redshift Survey using the Sloan Digital Sky Survey: Early Data Release
PASJ submitted – Hiroyuki Yoshiguchi, et al.

Stellar-Mass Black Holes in the Solar Neighborhood
ApJ submitted – James Chisholm, et al.

The Pairwise Velocity Distribution Function of Galaxies in the LCRS, 2dF, and SDSS Redshift Surveys
ApJL accepted – Stephen D. Landy

The Shapes of Galaxies in the Sloan Digital Sky Survey
AJ accepted – S. M. Khairul Alam, et al.

Q3 2002

Average spectra of massive galaxies in the SDSS
ApJ submitted - Daniel J. Eisenstein, et al.

Two Rare Magnetic Cataclysmic Variables with Extreme Cyclotron Features Identified in the Sloan Digital Sky Survey
ApJ accepted - Paula Szkody, et al.

An Overview of the Broad-band Optical Properties of Low-redshift Galaxies
ApJ submitted - Michael R. Blanton, et al.

Selection of Metal-poor Giant Stars Using the Sloan Digital Sky Survey Photometric System
ApJ Letters submitted - Amina Helmi, et al.

Three--Dimensional Genus Statistics of Galaxies in the SDSS Early Data Release
PASJ accepted - Chiaki Hikage, et al.

A First Look at White Dwarf - M Dwarf Pairs in the Sloan Digital Sky Survey
AJ submitted - Sean Raymond, et al.

Based on public SDSS data

A Survey for Large Separaation Lensed FIRST Quasars, II. Magnification Bias and Redshift Distribution
MNRAS accepted – Eran O. Ofek, et al.

Detection of Weak Gravitational Lensing magnification from Galaxy-QSO Cross-Correlation in the SDSS
ApJ submitted – Enrique Gaztanaga

Apparent Clustering of Intermediate-Redshift Galaxies as a Probe of Dark Energy
PRL submitted – Takahiko Matsubara

Narrow-line Seyfert 1 Galaxies from the Sloan Digital Sky Survey Early Data Release
AJ accepted - R. Williams, et al.

Magnetic White Dwarfs in the Early Data Release of the Sloan Digital Sky Survey
A&A accepted – B. Gaensicke, et al.

Broad Absorption Line Quasars in the Early Data Release from the Sloan Digital Sky Survey
ApJ Lett accepted – Alin Tolea, et al.

Galaxy Clustering in the Sloan Digital Sky Survey (SDSS): A First Comparision with the APM Galaxy Survey
MRNAS, 33:L21 (2002) – E. Gaztanaga, et al.

SDSS J124602.54+011318.8: A Highly Variable Active Galactic Nucleus, Not an Orphan Gamma-Ray Burst Afterglow
PASP, 114:587 (2002) – Gal-Yam, et al.

Appendix C Thesis Topics

Student	Email	Advisor
Daniel Harbeck	(dharbeck@mpia-hd.mpg.de)	Eva K. Grebel

Modelling of the SDSS photometric system through synthetic colors and application to stellar population studies

Vandana Desai	(desai@astro.washington.edu)	Julianne Dalcanton

An Empirical Measure of Galaxy Evolution in Clusters

Beth Willman	(willman@astro.washington.edu)	Julianne Dalcanton

A Survey for Resolved LSB Stellar Populations in the Local Group

Bart Pindor	(pindor@astro.princeton.edu)	Ed Turner

Identifying Strongly-Lensed Quasars in the SDSS Imaging Data

Tomo Goto	(ss96068@mail.ecc.u-tokyo.ac.jp)	Maki Sekiguchi

Finding galaxy clusters in Sloan data using color cut and enhancement method.

Lei Hao	(haol@astro.princeton.edu)	Michael Strauss

Emission-line properties of galaxies: large-scale structure, active nuclei, and metallicity.

Liam O'Connell	(l.a.o-connell@sussex.ac.uk)	Jon Loveday

Constraining Cosmological Models with SDSS

Katsuko Nakahira	(nakahira@a.phys.nagoya-u.ac.jp)	Satoru Ikeuchi

Candidate Selection for Moderately Distant Clusters of Galaxies Using Matched Filter Method

Randall R. Rojas	(rrojas@mercury.physics.drexel.edu)	Michael S. Vogeley

Voids and Void Galaxies in the SDSS

Brice Ménard	(menard@MPA-Garching.mpg.de)	Matthias Bartelmann

Angular Cross-Correlations between Background Quasars and Foreground Galaxies

Shiyin Shen	(shen@mpa-garching.mpg.de)	Wolfgang Voges & Houjun Mo

Statistical analyses of X-ray emissions from SDSS AGNs and galaxies

Brigitte Koenig	(bkoenig@mpe.mpg.de)	Ralph Neuhaeuser & Wolfgang Voges

Search for post-T Tauri stars and young Zero-Age Main-Sequence stars among SDSS stars

Tara Murphy	(tm@roe.ac.uk)	Avery Meiksin

Spectral Properties of Galaxies

Iskra Strateva (iskra@astro.princeton.edu) Michael Strauss
Double-Peaked Broad Emission Lines and the Geometry of Accretion in AGN

Christy Tremonti (cat@pha.jhu.edu) Tim Heckman & Guinevere Kauffmann
The Nature of Star Forming Galaxies in the SDSS

Yeong-Shang Loh (yeongloh@astro.princeton.edu) Michael Strauss
The evolution of Brightest Cluster Galaxies from SDSS data

Gajus Miknaitis (gm@astro.washington.edu) Christopher Stubbs
Constraining Cosmological Parameters with Type Ia Supernovae

Brian Wilhite (wilhite@oddjob.uchicago.edu) Richard Kron
Spectral Variability of Quasars in the SDSS

Nicholas M Ball (N.M.Ball@sussex.ac.uk) Jon Loveday
Parametric Descriptions of the Galaxy Distribution

Stefano Zibetti (zibetti@mpa-garching.mpg.de) Simon D.M. White
Low surface brightness features of galaxies and diffuse intracluster light detection

Paola Popesso (popesso@mpe.mpg.de) Hans Boehringer
Correlation study of the properties of the galaxy population and the X-ray emission in cluster of galaxies in the SDSS

Craig Wiegert (c-wiegert@uchicago.edu) Josh Frieman
Constraining Compact Dark Matter with QSO Equivalent Widths

Zsuzsanna Gyory (gyory@complex.elte.hu) Istvan Csabai
Photometric redshift estimation with model galaxy spectral templates

Jakob Walcher (walcher@mpia.de) Hans-Walter Rix
The local black hole mass function

Markus B. Huber (mhuber@mpe.mpg.de) Hans Boehringer
Studying the topology of the large-scale structure with new techniques

Nikhil Padmanabhan (npadmana@princeton.edu) Uros Seljak
Galaxy correlations as a function of stellar mass

Luigi C. Gallo (lgallo@mpe.mpg.de) Thomas Boller &, Wolfgang Voges
An X-ray-Optical Study of Narrow and Broad Line AGN with ROSAT and SDSS Data