The Future of the Sloan Digital Sky Survey: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe

1. Introduction

A great discovery of modern astronomy is that the origins of our planet, our Sun and the Milky Way are intricately tied to the nature of the early Universe and the laws of physics. The goal of the next phase of the Sloan Digital Sky Survey (SDSS) is to extend our understanding of these ties using a uniquely powerful astronomical facility, and to enhance the public's understanding of these fundamental aspects of our Universe.

In the 1920s, Edwin Hubble demonstrated that the Universe is expanding. Galaxies move away from us with a speed proportional to their distance, implying that they all emanated from a single point 13.8 billion years ago, at the "Big Bang." A decade after Hubble's discovery, Fritz Zwicky found the first evidence that most of the gravitating mass in the Universe is invisible. Over the following sixty years, evidence mounted that this mass was in some exotic form — not protons or neutrons or any other known particles, but something else now known as "dark matter." In the 1990s, astronomers made the more profound discovery that the Universe's expansion was accelerating, defying the expectation from Einstein's theory of gravity that its mass would cause it to slow down. This effect has given rise to a number of speculative models known collectively as "dark energy."

Dark matter and dark energy dominate the Universe and remain the greatest puzzles in physics, but we detect them only indirectly through astronomical observations of ordinary matter. Both dark components shape the large-scale expansion of the Universe, while on smaller scales tiny initial fluctuations grow into mammoth dark matter "halos," where galaxies like our Milky Way form. Inside galaxies, normal matter aggregates to form supermassive black holes and stars, generating the light we chart the cosmos with and the basic elements, like oxygen, nitrogen and carbon, from which planets and life arise. This interplay between gas, stars, black holes, dark matter and dark energy has only begun to be understood.

Over the past fourteen years, a consortium of universities and international partners, along with the National Science Foundation, the Department of Energy, and the Sloan Foundation, have executed three Sloan Digital Sky Survey projects (SDSS-I, -II, and -III). These projects created an extraordinary legacy of mapping structure across a vast range of scales, from asteroids in our own Solar System to quasars over 10 billion light-years away. The next phase of the SDSS will expand this legacy by studying the detailed assembly history of the Milky Way and thousands of other nearby galaxies, and extending precision measurements of expansion to the cosmic epochs when dark energy first became important. Three surveys (APOGEE, MaNGA and eBOSS) will use the Sloan Foundation 2.5-m Telescope at Apache Point Observatory (APO) in New Mexico from July 2014 to July 2020. The SDSS facility's key advantage is its wide-field spectrographs that observe many objects simultaneously, spreading the light of each into its separate wavelengths, yielding a *spectrum* that allows the study of each object's detailed physical nature. These spectra also yield Doppler velocities, which measure stellar motions, the internal motions within other galaxies, and galaxy distances based on Hubble's Law (see Figure 1).

The APO Galactic Evolution Experiment 2 (APOGEE-2) will explore the formation of the Milky Way using a comprehensive archaeological record based on 500,000 spectra of stars, revealing

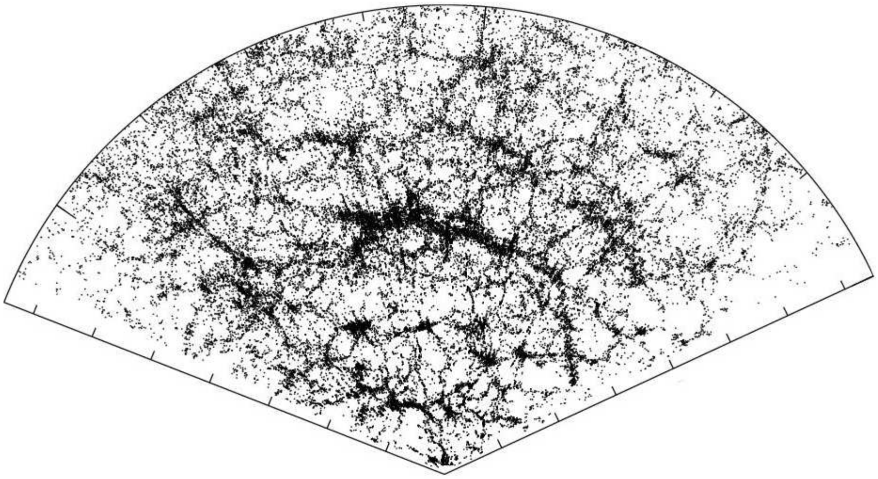

Fig. 1.— Map of galaxies from SDSS out to 1.5 billion light-years. The points are galaxies, with distances determined using Hubble's Law and Doppler velocities from spectra. The Milky Way is at the bottom and the wedge shown is one slice through SDSS's 3D map.

the structure of the Galaxy and its stars. Simultaneously, it will be able to detect "exoplanets" around other stars and in addition conduct an unprecedentedly large study of known exoplanet systems. The northern component (APOGEE-2N) will continue observations from APO while the southern component (APOGEE-2S) will observe the other half of the sky from the 2.5-m du Pont Telescope at Las Campanas Observatory (LCO) in Chile.

Mapping Nearby Galaxies at APO (MaNGA) will uncover the internal structure and formation history of 10,000 galaxies and characterize the diversity of their evolutionary histories. It will contain the largest sample of galaxies ever observed in resolved spectroscopy by more than a factor of ten. With many spectra across the extent of each galaxy to trace its assembly history and dark matter content, MaNGA will provide a uniquely rich legacy data set.

The Extended Baryon Oscillation Spectroscopic Survey (eBOSS) will achieve the best ever measurements of cosmic expansion to a distance of 12 billion light-years. It will provide the first measurements across the critical epoch between 6.5 and 11 billion light-years, the predicted "onset time" for dark energy. eBOSS will also include a massive sample of variable stars selected from time-domain imaging surveys, follow-up on a unique sample of X-ray sources, and create the largest existing sample of accreting supermassive black holes.

These surveys will address all the major themes of the National Academy's 2010 Decadal Survey of Astronomy & Astrophysics, advancing our understanding of fundamental physics, cosmology and the formation of galaxies, stars, and planets. This ambitious program is made possible by leveraging the existing SDSS expertise and infrastructure, and implementing low-risk but high-reward improvements. All data will continue to be distributed in a public archive using the successful and popular tools of previous SDSS surveys.

2. The SDSS Legacy

The SDSS has left an imprint on the science and culture of astronomy, transforming the study of the cosmos from the solar system to the most distant quasars. It has progressed in three distinct phases: SDSS-I (2000—2005), SDSS-II (2005–2008), and SDSS-III (2008–2014). The collaborating scientists have written over 1000 papers based on SDSS data, and including those written using our public data releases there are a total of 5000 SDSS-based papers with over 200,000 citations. Within the Collaboration there have been over 120 SDSS-based PhD theses, and outside the Collaboration there have been many more. The SDSS has several times been named the highest impact project, facility or mission in the field of astronomy, as judged by number of citations of associated refereed journal articles (Madrid & Macchetto 2009); the SDSS was the source of the most highly cited astronomy article in the years 2000, 2002, 2005, and 2008 (Frogel 2010).

The publicly available, user-friendly SDSS tools have fueled a large number of undergraduate and even high-school projects and research papers (including Siemens Competition winners). Images from the SDSS are the basis of Google Sky and Microsoft's World Wide Telescope, the two most used public interfaces to deep sky imaging. The Galaxy Zoo project based on SDSS is a hugely successful "citizen science" experiment that has involved over 100,000 members of the public in astronomical discovery, and that inspired the creation of the Citizen Science Alliance, which has now expanded the Galaxy Zoo framework to engage the public in a range of scientific fields.

The next phase of the SDSS will build on four themes representing our signature achievements:

Mapping the present-day galaxies. With a million high quality images and spectra, SDSS transformed our understanding of nearby galaxies by thoroughly characterizing the physical differences between spiral and elliptical galaxies and showing that transitions between these classes must be rapid. MaNGA's spatially resolved spectroscopy will expand the detail of this characterization of galaxies dramatically, providing a legacy data set of unparalleled comprehensiveness.

Establishing the standard cosmological model with precision measurements. The huge size and superb calibration of the SDSS galaxy surveys enabled unprecedent measurements of clustering and gravitational lensing, central pillars supporting dark matter and dark energy. Most strikingly, SDSS definitively detected the baryon acoustic oscillations (BAO). In SDSS-III, the BOSS survey uses the BAO to determine the most accurate absolute distances available in cosmology today. eBOSS will extend these studies into more distant epochs of cosmic history, seeking clues about the nature of dark energy.

Structure and formation of the Milky Way Galaxy. SDSS star maps demonstrated the ubiquity of stellar streams and in-falling dwarf galaxies in the outer Galaxy, suggesting a complex assembly history punctuated by multiple and ongoing "cannibalization" of smaller galaxies (see Figure 2). In SDSS-III, APOGEE is revolutionizing Milky Way science yet again, with exquisitely precise measurements of the histories of individual stars, and a unique ability to penetrate the dust-obscured inner Galaxy. The APOGEE-2 survey will quintuple our sample size, and its southern hemisphere component will map the heretofore poorly known but essential central regions at the heart of the Milky Way.

A comprehensive census of active supermassive black holes. The most luminous objects in the Universe are quasars, which are supermassive black holes at the centers of galaxies whose gravity energizes hot, glowing gas falling into them. With more than 200,000 quasar spectra, the SDSS mapped the growth history of black holes back to epochs less than a billion years after

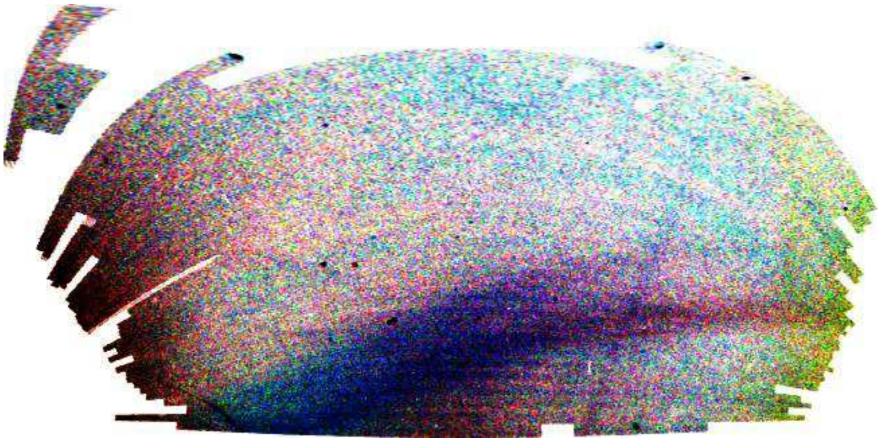

Fig. 2.— Distribution of stars in SDSS imaging data in the northern Milky Way stellar halo (courtesy V. Belokurov). Color denotes distance (blue is further than red). The image is pockmarked by small star clusters and dwarf galaxies, but dominated by several major and minor star "streams" showing that the Milky Way is experiencing the ongoing infall and disruption of numerous small satellite galaxies.

the Big Bang. eBOSS will more than double this sample and use new techniques to create the most complete census of quasars to date, shedding new light on their growth and their interaction with their host galaxies.

With a wide-field spectroscopic capability that far surpasses any other existing facility, the SDSS facilities and software infrastructure are intact and available for the efficient implementation of a follow-on project. Through each phase, the SDSS collaboration has proven robust and flexible and provides a strong basis for the future. We will spend the coming years exploiting and enhancing these resources, with the same focus on high-impact, high-quality, open-access science.

3. Extending APOGEE: A Comprehensive Unveiling of the Milky Way

APOGEE-2 will unravel the history of the Milky Way, both as our home in the Universe and as a representative massive spiral galaxy. Simultaneously, it will find a host of stellar companions (including exoplanets) and provide unique follow-up observations for NASA's Kepler satellite to study stellar ages and the relationship between stellar chemistry and planet formation. It will use the APOGEE near-infrared spectrograph to precisely measure 500,000 individual stars. This comprehensive sample will allow a detailed "archaeological" reconstruction of the formation of our Galaxy. APOGEE-2N will continue observations of the outer Galaxy at APO. APOGEE-2S will complete the study of the inner Galaxy, mainly visible only in the southern sky, from the du Pont Telescope in Chile (run by the Carnegie Institution for Science).

Like many galaxies, the Milky Way has an old central region (the "bulge") surrounded by a massive disk of gas and younger stars, and a "stellar halo" of ancient stars reaching to large distance. These stars and gas sit near the center of a much more massive and extended dark matter halo. Studying all these components is crucial to determining the Milky Way's history. The outer disk is where the Galaxy primarily grows today, whereas the inner disk and central region are the sites of its historical growth. The Galactic stellar halo has strong substructure (Figures 2 and 3) revealing how satellite infall helped to build the Milky Way.

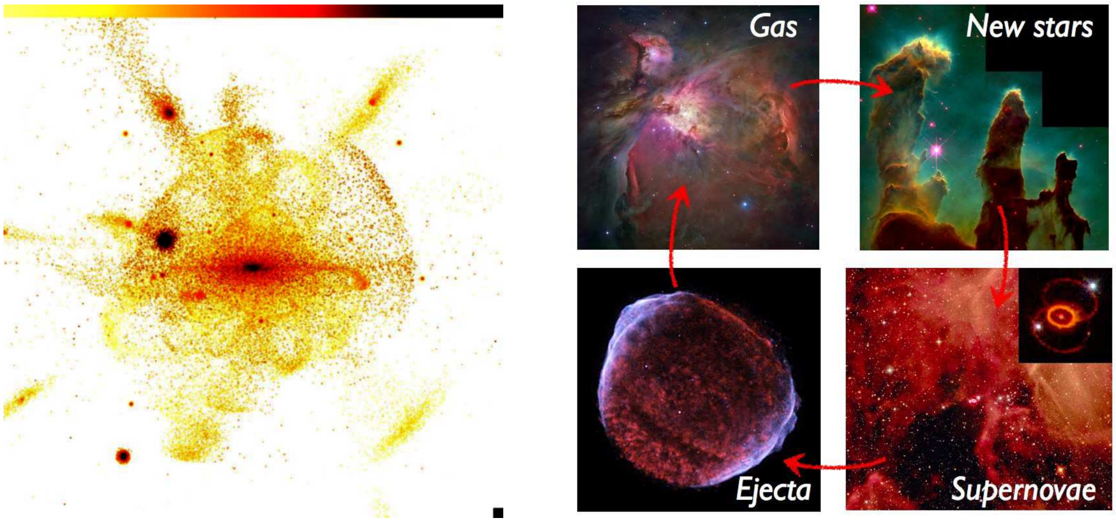

Fig. 3.— *Left:* Theoretical model of the structure of a Milky Way-like galaxy, showing the results of growth through the infall and destruction of smaller galaxies (image courtesy J. Bullock). These “streams” are observed in the real SDSS data and shown in Figure 2. Understanding their elemental abundances and dynamics will yield a detailed understanding of their history. *Right:* Life cycle of stars: Gas collapses gravitationally to form stars, which produce new elements, which are returned to the gas supply through winds and explosions, and a new, “enriched” generation of stars is born. APOGEE measures this enrichment history to “chemically tag” stellar siblings and unravel the history of the Milky Way.

From our vantage point inside it, we can understand the Milky Way better than any other galaxy by studying it star by star. In the cores of stars, primordial hydrogen and helium formed during the Big Bang are fused into higher mass elements, like carbon, nitrogen and oxygen. These new atoms are returned by stellar winds and explosions to the Milky Way’s disk, where the next generation of “enriched” stars form. Meanwhile, fresh unenriched gas is flowing into the Galaxy from the outside. As illustrated in Figure 3, this complex history of star-formation and gas recycling is encoded in the observed patterns of higher mass elements. By measuring these “fossil” patterns in stars across the entire Milky Way, APOGEE-2 will determine in detail its formation history in the deep past — a form of Galactic archaeology.

The APOGEE-2 survey will use a 300-fiber spectrograph built for SDSS-III that operates at near-infrared wavelengths three times longer than visible light. The wavelength band contains myriad spectral features revealing the presence of iron, oxygen, carbon, nitrogen and numerous other elements. The spectra from this instrument provide accurate determinations of the abundances of these elements and extremely precise Doppler velocities of stars, good enough to detect the tugs of planets circling the observed stars. APOGEE in SDSS-III has already observed 50,000 stars and is the largest data set of its kind by a factor of 100. Existing software pipelines determine abundances and Doppler velocities automatically. The first scientific results from APOGEE are now being submitted to the journals. APOGEE-2 will extend the original APOGEE survey to a thorough and unprecedented spectroscopic assessment of the entire Galaxy in both hemispheres.

There is no comparable competitor program to APOGEE-2 in existence at any telescope. APOGEE-2 wields three major advantages for studying the entire Milky Way. First, the Galactic bulge and disk are heavily obscured by interstellar dust, which to date has stymied their detailed

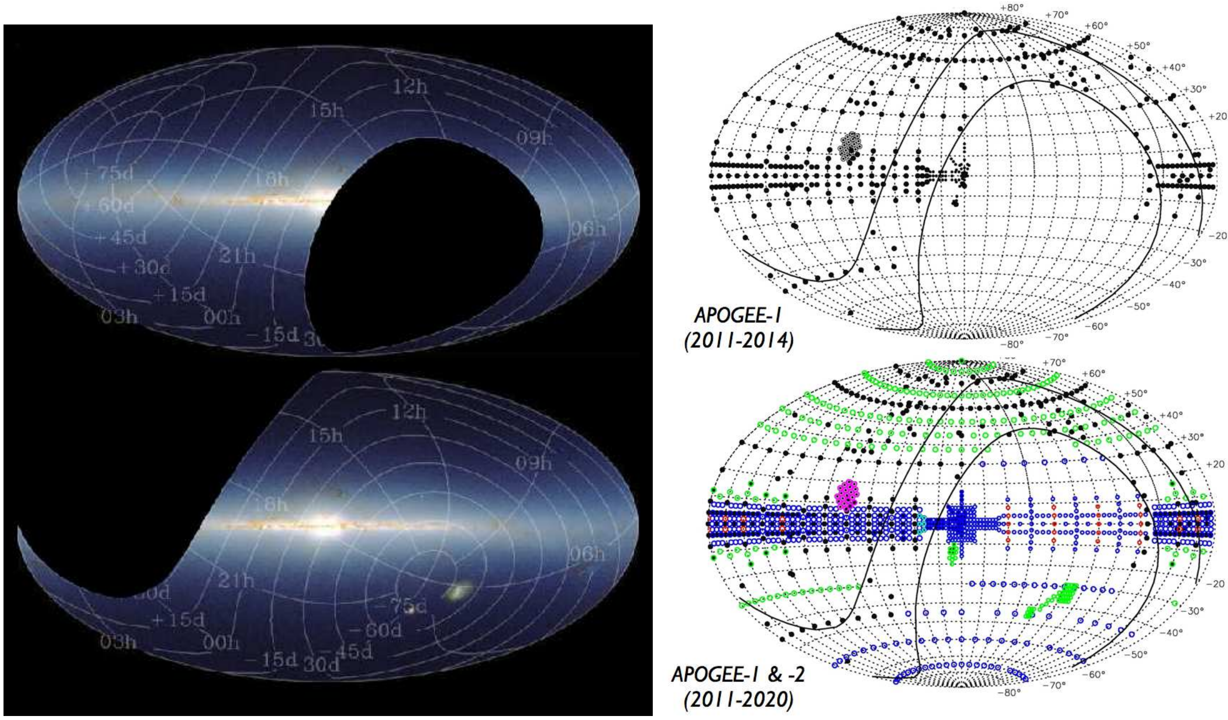

Fig. 4.— *Left:* Panoramic views of the accessible sky at the Sloan Foundation (*top*) and the du Pont (*bottom*) Telescopes. The 2MASS data in these images prominently show the Milky Way structure and the Magellanic Clouds. *Right:* The APOGEE targeting plan for SDSS-III (*top*) and the next phase of SDSS (*bottom*).

study. APOGEE's observations peer deep into this veil of dust, because at near-infrared wavelengths the dust obscuration is far lower than at visible wavelengths. Second, to study the inner and outer Galaxy and the full structure of the stellar halo requires access to the entire sky. APOGEE-2's southern component at the du Pont Telescope delivers this capability (Figure 4). Third, the near-infrared allows the best measurement of the most abundant elements in the Universe, such as carbon, nitrogen and oxygen. Combined, these advantages allow APOGEE-2 to study all regions of the Milky Way with a massive, homogeneous, detailed spectroscopic survey for the first time.

The first two advantages mirror those of the Two-Micron All Sky Survey (2MASS), which imaged the entire sky in the near-infrared from both southern and northern observatories, producing a huge catalog of stellar photometry and generating clearer views of the Galaxy (Figure 4). APOGEE-2 will provide the first comprehensive spectroscopic follow-up survey for 2MASS, yielding a unique long-term legacy value.

APOGEE-2's quintupled sample size will enable new science greatly exceeding that of APOGEE-1 alone, allowing us to fully map the stellar populations left over from the formation of the Galactic halo and disk. In the stellar halo, APOGEE-2N will much more comprehensively characterize the chemistry of substructure, shedding light on how the Milky Way halo was assembled. In the disk, the increased sample size will allow us to identify *every* accreted star-forming site. The accurate kinematic measurements in the disk and halo will enhance measurements of the Milky Way dark matter halo structure. Moreover, APOGEE-2N will extend the temporal baseline of our spectroscopic searches for planetary and other substellar companions such as neutron stars and white dwarfs, the progenitors of supernovae. APOGEE-2N will comprise the largest ever spectroscopic search for these companions.

Furthermore, APOGEE-2N will greatly enhance a fruitful synergy with NASA's Kepler mission, which images a northern field continuously to find planets around other stars and measure stellar pulsations ("asteroseismology"). APOGEE-2N will provide accurate chemical measurements of both planet hosts and non-hosts to understand better the role that key elements — such as carbon, nitrogen, oxygen, magnesium, silicon and iron — play in the formation of both gaseous and rocky planets and investigate how these elements influence the potential habitability of the latter.

Moreover, combining APOGEE-2N data and Kepler asteroseismology will yield high precision measurements of the ages of individual stars. Stellar age is extremely difficult to measure and is a precious piece of information for understanding galaxy evolution. A pilot program in SDSS-III demonstrated the enormous power of combining APOGEE and Kepler, and based on this success APOGEE-2N will devote a full-scale campaign to the Kepler fields to attain a comprehensive, order of magnitude larger census of Kepler targets. This is a truly unique opportunity, since there are no other high throughput, high resolution spectrographs in the world that can access the Kepler field.

For APOGEE-2S, we plan to move the APOGEE instrument in January 2017 to the 2.5-m du Pont Telescope in Chile, and observe the remaining years in the Southern hemisphere. An enhanced goal (depending on funding) is to build a second instrument and run the two surveys simultaneously, greatly increasing the scientific yield of both. In either case, we will apply the modifications necessary to the du Pont observing systems to run in the efficient manner of the SDSS.

From the south, APOGEE-2S has ideal views of the far side of the disk plus the older components of the Galaxy essential for understanding its ancient past — its bulge, bar, inner disk, and the globular clusters. Globular clusters are the oldest stellar systems known and essential objects for calibrating our archaelogical interpretation of the Galaxy. The southern sky also holds the three major nearby dwarf galaxies falling into the Milky Way (Sagittarius and the two Magellanic Clouds).

APOGEE-2 will be a unique legacy data set — a treasure trove of Galactic knowledge that will take years to mine. Furthermore, with its full coverage of the disk and stellar halo APOGEE will become a touchstone and common denominator for all future experiments to come. APOGEE is already revolutionizing stellar spectroscopy. Previous to APOGEE, astronomers analyzed the existing, much smaller samples one-by-one, occupying hours of human time per star. APOGEE is incorporating the collective wisdom of these stellar spectroscopists and incorporating it into innovative computational techniques for automatic, efficient and homogeneous analysis. This success repeats that of previous SDSS surveys, which similarly transformed the interpretation of visible wavelength spectra. APOGEE-2 will establish a new standard in quality and reliability for stellar spectroscopy in the service of learning about our own Galaxy's formation and fundamental stellar astrophysics.

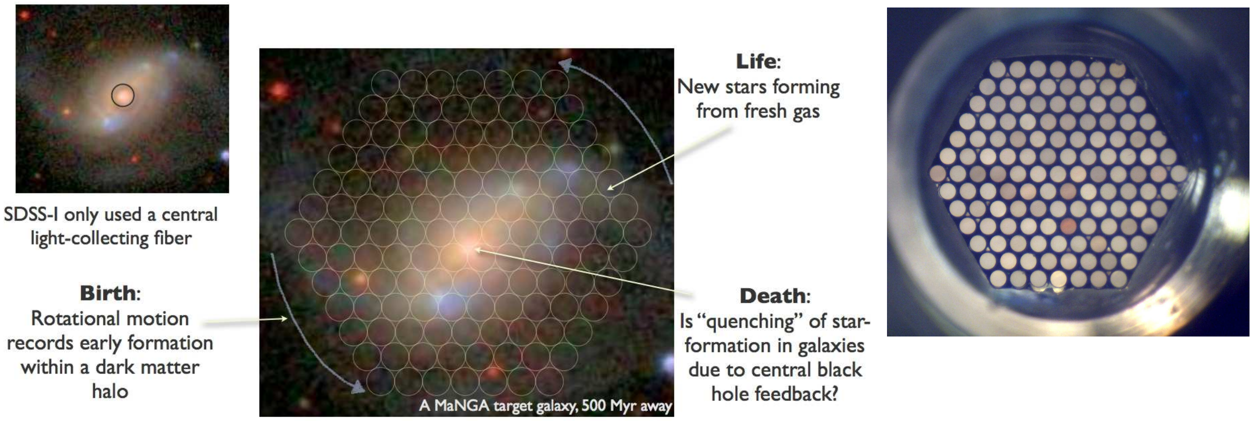

Fig. 5.— A schematic description of MaNGA's ability to probe fundamental questions in our understanding of galaxy formation and evolution. *Left:* Coverage of a standard single-fiber observation of a nearby galaxy. *Middle:* MaNGA resolved spectroscopic coverage of the same galaxy. The grid of circles represents the 127 fibers in MaNGA's largest fiber bundles. *Right:* Example of the actual hex-packed fiber bundles recently built and tested on-sky by the MaNGA team.

4. MaNGA: dissecting the nearby galaxy population

MaNGA focuses on understanding the variations in the galaxy assembly process by comprehensively measuring a large galaxy sample. Whereas most surveys of distant galaxies obtain a single spectrum at the center of each, MaNGA will measure up to 381 independent spectra across the face of each of 10,000 galaxies. For studying nearby galaxies, this survey in "resolved spectroscopy" will provide the richest legacy data set ever created. In particular, these maps will tell us how and when galaxies were formed and the nature of their continued growth and evolution. They will tell us how much the life cycles of galaxies are governed by their own basic properties ("nature") versus the environments they live in ("nurture").

MaNGA will achieve these breakthroughs by combining the technology behind the successful SDSS surveys with an innovative twist that enables resolved spectroscopic measurements of every galaxy. As shown in Figure 5, instead of measuring one spectrum at the galaxy's center, MaNGA will deploy fiber bundle arrays, sampling each target galaxy at up to 381 individual locations (depending on the galaxy's apparent size). These fibers will be connected to the existing BOSS spectrograph and exploit the well-understood quality and performance of that instrument. With the large available field of view, MaNGA can deploy such fiber-bundles on 20 separate galaxies simultaneously. Previous instruments capable of resolved spectroscopy typically targeted only one galaxy at a time. MaNGA therefore enjoys a "multiplex" advantage and can quickly amass very large samples.

The resulting statistical power allows us to better explore the diversity of galaxies in the Universe. Over the age of the Universe gravity has amplified tiny fluctuations in the early Universe into the rich structure seen in galaxies today. The Milky Way is a typical disk galaxy, with spiral arms and ongoing star-formation, but many massive galaxies are ellipticals that are no longer forming stars (see Figure 6). Spiral galaxies are found preferentially in empty regions of the Universe, but elliptical galaxies are found mostly in the densest parts, surrounded by up to hundreds of neighbor galaxies. There is a large variety within each of these types, and their growth is regulated by the accretion of new fuel for star-formation, internal “feedback” halting star-formation due to supernovae and accretion onto the central black hole, and occasionally violent galaxy mergers.

From SDSS-I and -II we have detailed images of nearly a million nearby galaxies but spectra of just their central regions; this only yields partial information on their history. The life stories of these galaxies are encoded in their internal structure — as the life story of a tree is encoded in its rings — and only spectra can reveal this structure. MaNGA will map out this story by measuring the fundamental galactic building blocks: the dark matter whose gravity binds the galaxy, the gas from which stars form, the stars themselves, and the chemical elements that these stars produce in their nuclear furnaces and then return to the galaxy during explosive deaths. We will be able to map out both the history of the formation of stars and the bulk motions of the stars and gas at each location in each galaxy.

MaNGA’s array of spectroscopic measurements across the face of each galaxy will address fundamental questions regarding the birth, life, and death of galaxies (see Figure 5). Spectroscopy reveals the abundances of elements in the stars and in the gas that is forming new stars, as well as their motions within the galaxy. Most galaxies experience a long life, continuing to grow and form new stars from fresh gas supplies. Eventually, however, star formation is somehow shut down in many galaxies, a kind of “galaxy death” that may be the fate of our own Milky Way. The challenge for the next decade is to understand the physical processes that drive each stage in this process.

Birth. The formation of the dark matter halo and its interaction with the collapsing cloud of gas that eventually forms the proto-galaxy is encoded in its 3D velocity structure and preserved to the present day. MaNGA’s resolved spectroscopy will map this structure and allow us to understand how stars and gas interact with their host dark matter halo and at what point major components of the galaxy like its central bulge and disk formed.

Life. The majority of local galaxies are still growing, but we do not understand the origin of the fresh gas needed to fuel ongoing star formation, nor do we know how and at what rate mergers with other galaxies contribute to growth in galaxy size and mass. MaNGA will map where star formation is occuring, telling us whether the fuel is being delivered by external streams or is condensing like dew from the warm gas already present at the center. Our maps will reveal differences in composition between the center and outskirts, a key test for whether mergers build galaxies from the outside-in or inside-out.

Death. The last stages are particularly mysterious. Galaxies do not simply run out of fuel, but instead their star formation is quenched by a still unknown process. Is it heating from accretion onto a central supermassive black hole? Or perhaps a bath of hot plasma that prevents cold fuel supplies from reaching the starving galaxy? Maps of winds and shocks likely driven by this process, as well as the location in each galaxy where star formation first seems to end, will allow MaNGA to provide definitive answers.

MaNGA’s technical improvements on the current system are low risk. It takes full advantage of the high sensitivity and well studied performance of the BOSS spectographs and the SDSS observing system. The biggest technological challenge is the development of the fiber bundle arrays. The MaNGA team includes members from several universities with long experience in this technology, which has now been fully demonstrated. Eight prototype bundles from a proven vendor were successfully tested on sky at Apache Point in December 2012. The key task now is to produce these bundles cost-effectively in bulk.

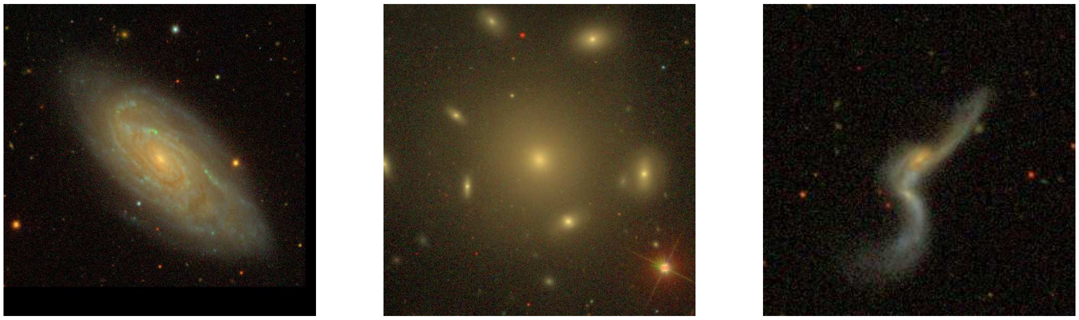

Fig. 6.— Galaxies in SDSS imaging. *(Left)*: a typical spiral galaxy. *(Middle)*: a typical elliptical galaxy (surrounded by many neighbors). *(Right)*: a merger of galaxies.

MaNGA will surpass any of its potential competitors. Ten years ago, resolved spectroscopy was pioneered by the highly successful SAURON survey of 50 galaxies, which revealed that otherwise similar-looking elliptical galaxies could have radically different internal structures. The ATLAS^{3D} project has now expanded that sample to 260 and the CALIFA survey has begun, which plans to observe 600, still more than a factor of ten fewer than MaNGA. The closest competitor to MaNGA is the SAMI project at the Anglo-Australian Telescope (AAT), which observes the southern sky and might attain a sample half the size of MaNGA's, if granted adequate time on the AAT. In previous years, the AAT and SDSS projects have often productively worked in parallel on similar science objectives, focusing on different regions of sky. However, the BOSS spectrographs used by MaNGA yield a larger wavelength range with better velocity resolution. SDSS has also distinguished itself by committing to innovative public data releases that have become a reliable legacy data resource. MaNGA will distribute both the largest and the most user-friendly survey of resolved spectroscopy.

The field of astronomy transformed completely when moving from small, disparate galaxy surveys to the panoramic sample of the SDSS. MaNGA will provide a similar quantum leap with homogeneus resolved spectroscopy for more than ten times the number of galaxies currently available. It will provide a necessary complement to studies of young galaxies in the distant Universe, the focus of the future generations of major new facilities.

Many key members of previous surveys of resolved spectroscopy and of BOSS are now MaNGA team members. Capitalizing on their experience, we will develop new techniques to automatically, homogeneously and optimally interpret these spectra. As in the case of APOGEE-2, the spatially resolved spectroscopy of MaNGA represents an opportunity to apply for the first time modern techniques to automatically and homogeneously extract the maximum amount of physical information. We will also develop a new set of access and vizualization tools to simultaneously model and interpret many spectra per object, constituting a legacy software resource all its own for interpreting future surveys.

5. eBOSS: Probing the Dark Side of the Universe

eBOSS focuses on measuring the expansion of the Universe over its entire history and understanding what it reveals about fundamental physics, particularly regarding dark matter and dark energy. To do so, it will use the BOSS spectrograph to create a map of the Universe spanning a wider range of distances than any other previously obtained.

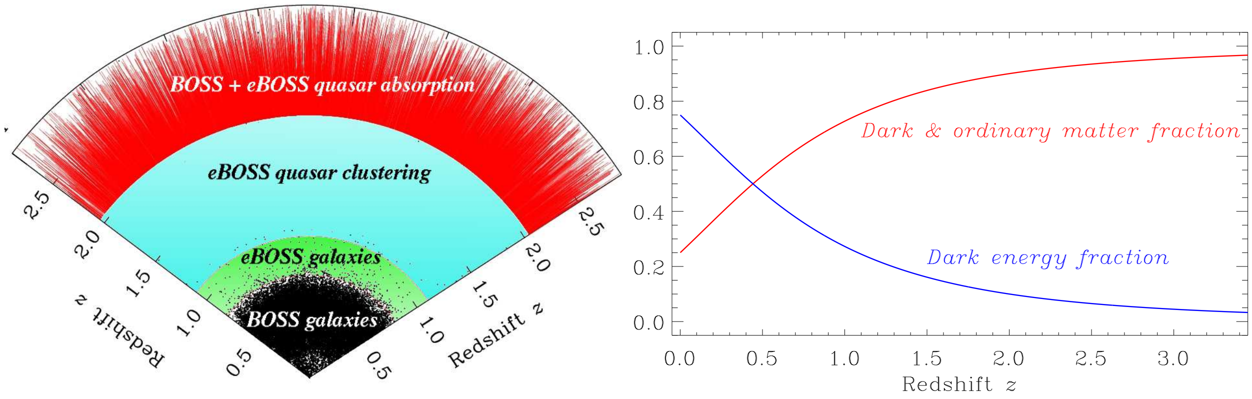

Fig. 7.— *Left:* eBOSS redshift coverage. eBOSS will be the first large-scale structure survey in the critical range $0.7 < z < 2$. *Right:* Fraction of energy density due to dark and ordinary matter (red line) and dark energy (blue line), showing onset of dark energy at $z \sim 2$.

Over the past century, we have unearthed problems in fundamental physics that can only be studied using astronomy. Hubble's first measurements of the relation between velocity and distance of galaxies revealed the expansion of the Universe. Today we know that the expansion is accelerating and is regulated by dark matter and dark energy, which comprise 96% of the Universe's energy density. The Nobel Prize in Physics in 2011 was granted for the definitive discovery in 1998 of this acceleration.

In the last decade, numerous theories have been advanced to explain the accelerated expansion, but its true nature is unknown. The goal of cosmologists is to precisely measure the contents of the Universe and to understand their behavior over time. Doing so requires understanding the onset of the era of dark energy around 6.5–11 billion years ago — the time that dark energy's influence over the Universe's dynamics started to become important.

This challenge motivates us to design a new survey with a larger reach. Its primary goal remains to measure the Hubble Law relating the distances of objects to their recession velocities. More distant objects recede more quickly, and because their light takes longer to reach us, exactly *how much more quickly* reveals the history of the Universe's expansion. In measuring the Hubble Law, the velocity is easy to determine from Doppler shifts and is quantified as a redshift z, ranging from $z \sim 0$ nearby to $z \sim 3$ at a distance of 12 billion light-years away. However, the exact distance is extremely difficult to measure.

SDSS has made fundamental contributions to studying the Hubble Law. Its signature accomplishment in this arena has been to establish the utility of the "Baryon Acoustic Oscillation" (BAO), the most accurate known measurement of absolute cosmic distance. For the first 400,000 years, the Universe was filled with a tightly coupled gas consisting of matter and light. Small perturbations drove pressure waves through this gas. As the Universe expanded, the matter became less dense and more transparent to the light, whereupon these waves stalled. Today, the remnant of these waves produces small features (the BAO) in the very large scale ($\sim$ 500 million light-years) clustering of galaxies. The BAO is at a known physical scale, and can be detected in galaxy redshift surveys, so it can be used to measure the absolute physical distance to a particular redshift. Because the physics of the BAO is simple, there are few systematic uncertainties in this measurement, making it the most accurate cosmological distance indicator known to date.

SDSS-III includes the Baryonic Oscillation Spectroscopic Survey (BOSS), which is measuring the BAO feature to high precision. By mid-2014, BOSS will map one quarter of the entire sky, measuring spectra for nearly 1.4 million luminous red galaxies to redshift $z < 0.65$ and over 160,000 quasars spanning redshift $2.2 < z < 3.5$. During spring 2012, the BOSS collaboration announced the first high-precision measurement of the BAO peak from the galaxies at redshift $z \sim 0.57$, extending earlier SDSS results by powerfully constraining the relationship between distance and redshift out beyond the local Universe. In a major new result published in Fall 2012, BOSS has for the first time detected the BAO in the distribution of matter imprinted on the absorption spectra of quasars, yielding a few percent level measurement of the distance to $z \sim 2.5$, a leap past any previous cosmological distance measurement. However, neither BOSS nor any ongoing survey will probe the intermediate redshifts $z \sim 0.7$–2, the window in which dark energy starts to become important. Extending BAO measurements to this redshift window is thus highly desirable.

eBOSS will study dark matter and dark energy by accurately constraining the expansion history of the Universe across all redshifts from $0.6 < z < 3.5$ over a footprint of 7500 sq. deg. Using proven techniques, we will measure BAO from 1.4 million targets in four different classes:

1. elliptical galaxies (375,000 in the range $0.6 < z < 0.8$);
2. star-forming galaxies (260,000 in the range $0.6 < z < 1.0$ over a 1500 sq. deg. subregion);
3. "mid-redshift" quasars (700,000 in the range $1 < z < 2.2$);
4. "high-redshift" quasars (40,000 in the range $2.2 < z < 3.5$).

The elliptical galaxies, star-forming galaxies, and mid-redshift quasars will show BAO in their clustering, and will provide the first BAO distance measurement in the range $1 < z < 2$. The high-redshift quasars will show the BAO signal in absorption and greatly enhance the BOSS sample. eBOSS can select all of these targets from SDSS imaging and other public data sets. Combined with previous SDSS results, the resulting distance constraints will be better than $\sim$ 2–3% accuracy covering $z \sim 0.6$ all the way to $z \sim 3.5$.

eBOSS's massive scale allows other important constraints. Geometrical constraints such as the BAO can be combined with probes of the growth of structure, such as redshift-space distortions (which tell us how galaxies are moving in response to gravity) and weak gravitational lensing (which tells us how massive objects bend the path of light). Both of these methods can not only measure dark energy but can independently test for deviations from Einstein's theory of gravity. Furthermore, the distribution of eBOSS targets on very large scales can constrain the mass of the neutrino and the nature of the "inflationary epoch" in the very early Universe, which produced the primordial fluctuations that seeded the formation of galaxies and larger structures.

Two eBOSS subprograms will target special classes of objects. First, the Time Domain Spectroscopic Survey (TDSS) will select time-variable targets for — in some cases repeated — spectroscopic follow-up. These targets will include quasars and several classes of variable stars, some of which may reveal previously unidentified phenomena. Major time-domain imaging surveys currently ongoing are finding numerous such targets, but their classification is stymied by the lack of a massive, unbiased follow-up spectroscopic program. By spectroscopically identifying the physical nature of $\sim 10^5$ detected variables, TDSS will greatly enhance the power of these existing data sets to illuminate the physics of stars and quasars. Future photometric surveys (such as the Large Synoptic Survey Telescope) will leverage TDSS observations to classify their detections more robustly.

Second, the SPectroscopic IDentification of ERosita Sources (SPIDERS) survey will take spectra of $\sim 5\times10^4$ X-ray emitting quasars and X-ray cluster galaxies detected by eROSITA, a groundbreaking new X-ray telescope to be launched in 2014. Quasars emit copiously in X-rays, but some are obscured in visible light. This telescope combined with eBOSS follow-up will uncover this hidden population of obscured quasars. Because X-ray observations also allow robust measurements of black hole growth rates, eBOSS will provide the most complete measurements to date of the growth history of supermassive black holes. Meanwhile, SPIDERS observations of X-ray cluster galaxies will allow us to map the 3D distribution of these X-ray clusters — the most massive bound systems in the Universe, exceeding 10^{15} times the mass of our Sun, and key tracers of the growth of structure.

Finally, the SDSS experience has shown that the serendipitous science impact of broad surveys far exceeds prior expectations. The eBOSS galaxies at $z \sim 1$ will explore the most extreme star-formers in the Universe's history and reveal rare phenomena such as outflows and strong gravitational lenses. With its broad sample of quasars, galaxies, X-ray, and variable sources, we can expect eBOSS to discover many previously unknown phenomena.

There is no other existing facility that can accomplish all the goals of eBOSS. HETDEX, a spectroscopic dark energy-focused survey just ramping up, will probe redshifts $z \sim 2$–3 using a complementary method. The Dark Energy Survey (DES) is a photometric survey that will address questions in cosmology using the BAO feature measured purely with imaging data. Relative to DES, the power of eBOSS is spectroscopic mapping in 3D, which greatly improves the cosmological precision. In fact, DES has recently altered its survey footprint to include a greater fraction of the planned eBOSS area, because spectroscopic data is crucial to capitalizing on its cosmological measurements.

Plans exist to build other facilities to address eBOSS science questions in the more distant future. The Department of Energy is planning the Mid-Scale Dark Energy Spectroscopic Instrument (MS-DESI) to start in 2018. The MS-DESI team and the eBOSS team are closely related, and MS-DESI's targeting strategy and spectroscopic reductions will rely heavily on the experience from eBOSS. The Prime Focus Spectrograph (PFS) is planned on the Subaru 8-m Telescope, with its first light also currently scheduled in 2018. While a cosmological program could be crafted using PFS, its science focus has not yet been determined. Neither of these projects has yet been completely funded, has begun the construction of its instruments or software, or has a well-defined public data release plan. eBOSS inherits a working telescope, instrument, and pipeline and will begin on schedule in Fall 2014. Furthermore, the targeting and data reduction strategies proposed for future surveys are patterned on eBOSS, which will provide critical experience necessary for future experiments if they do occur.

eBOSS will provide the first measurements of the Universe's expansion history and the growth of structure in a critical, yet uncharted, realm ($z \sim 0.7$–2.2), and in so doing study quasars, galaxies, clusters, lensing and large-scale structure. Combined, the three surveys APOGEE-2, MaNGA and eBOSS will span a vast range of cosmic time and scales, studying all aspects of the Universe from individual planets to its largest structures.

6. Science and Education with Astronomical Data

At the heart of the unprecedented impact of the SDSS is the data delivered by the SDSS telescope, instruments, and software systems, which represents the largest digital image of the sky and the largest three-dimensional map of our Universe. The project provides a real-world environment where we are developing and incorporating the latest tools to advance science with large, complex, and information-rich data sets.

A spectroscopic survey such as SDSS develops many layers of information that need to be processed, interpreted, archived, documented and distributed, with numerous interactions between the software and the physical systems executing the survey. First, targeting software analyzes images of the sky, choosing interesting spectroscopic targets and tracking selection effects for later use. Next, the spectroscopic observing plan is designed and machining code is written to execute the plate drilling. At the observatory, the plates are observed, and during the observing night real-time software systems control and log the observing progress and results. The raw data is transferred off-mountain, archived, and analyzed the next day. This analysis converts the raw spectroscopic images into individual calibrated spectra, redshifts, and derived physical parameters of the observed objects. The results are stored in documented flat-files and databases. On a yearly basis, the data is made available to all astronomers and the public through well-documented data releases, with tools suitable for a range of users, including astronomical experts as well as college and high school students. All the procedures and tasks are developed by federated yet interdependent teams of researchers within multiple collaboration working groups.

Using this data and its associated meta-data, the survey teams (and independent astronomers) perform the high-level science, such as measurements of how galaxies cluster, the expansion of the Universe, how galaxies grow, the elemental abundances within stars, and the structure of the Milky Way. To accurately interpret these results, which rely on subtle features measured in highly precise data, requires a careful analysis incorporating the data and meta-data of all of the above pipelines.

Providing a suitable product for such analyses requires the maintenance of high standards of *quality* and *accessibility.* To do so, SDSS has pioneered the application within astronomy of image and signal processing, machining learning and statistics, as well as the broad use of modern database and web tools for data distribution. We apply these techniques in all of the layers described above, including production, analysis and calibration of the data — developing, for example, machine learning techniques and network algorithms for target selection and survey optimization. The fundamental importance of all these tools that is now taken for granted in the astronomical survey community was first established by the SDSS.

Quality: SDSS set a new standard in astronomy for the development and application of advanced algorithms for imaging and spectroscopic data analysis to achieve images, spectra, and catalogs of uniformly high quality. This achievement has enabled new directions in quantitative science that would not have been otherwise possible, for example, detailed physical characterizations of the observed galaxies and stars that far exceeded the anticipated goals. The next-generation SDSS survey will continue this pattern of innovation, having assembled teams with world-leading expertise. All three core programs will incorporate new algorithms for spectroscopic data analysis that will continue to be at the forefront of astronomical practice. The resulting analysis pipelines will incorporate a new software library of data-management tools to facilitate uniform data handling practices across all three programs. These tools will provide easy programmatic access to all levels of the data and meta-data for the survey. They will allow the development of new analysis techniques to be built on top of or incorporated into the core pipelines.

Access: The SDSS data have been made available to the collaboration, to the wider astronomical community, and to the general public through regular planned and supported data releases. These publicly released data have been comprehensively documented, enabling users outside the SDSS collaboration to exploit the full scientific potential of the SDSS database, thereby multiplying its scientific productivity. SDSS data are now regarded as an indispensable public resource in astronomy. In a typical quarter, the SDSS SkyServer system logs nearly one million connections from the public, and over 300 individual astronomers regularly connect to our publicly available "advanced user" databases. For the next generation of SDSS, we will continue this practice with five public data releases: each summer from 2016 through 2019, and the end of 2020. The Science Archive Server will allow direct access to our detailed images and spectra, while the robust and powerful Catalog Archive Server (CAS) system will serve the catalogs based on them. SkyServer will remain the education and public outreach face. These tools will be upgraded to host the new data and new types of data from APOGEE-2, eBOSS and MaNGA. The same computing infrastructure will serve the collaboration and the public, ensuring both efficiency and robustness.

We will continue to facilitate "citizen science," such as the successful Galaxy Zoo and Galaxy Zoo 2 projects using the SDSS public data, by providing programming interfaces to all levels of our data, by partnering with the Citizen Science Alliance and other organizations, and by cooperating with amateur astronomers. There are particularly rich opportunities in MaNGA and APOGEE-2, for which we give just two examples. First, MaNGA galaxy velocity and abundance maps are rich and multi-dimensional; with 10,000 galaxies to explore, we anticipate an important opportunity for crowd-sourcing the examination of this population. Second, APOGEE-2 stars are commonly bright enough to be accessible to optical imaging through small telescopes, which can be used to providing time-domain monitoring for APOGEE-identified variable stars.

Furthermore, we plan to enhance the SkyServer's strong educational component, working in concert with educators to provide new educational projects for students and also new capabilities, such as an online "notebook" feature to help manage and evaluate student projects and a toolkit for accessing and analyzing data to lower the barrier to begin undergraduate research projects.

This suite of tools for science and education yields an ideal platform for efforts in data intensive sciences. The development of all the SDSS tools is science-driven and necessarily iterative. New science and education use cases arise over time that pose new algorithmic and technical challenges. The demanding requirements of scientists and their continuous, rigorous testing have been a necessary component to focusing the data analysis and management teams towards the most impactful problems. This close collaboration between scientists, computer scientists and software engineers is the best way to identify the subtle but important features that to enable science must be embedded deeply into the data and meta-data system — for example, the careful tracking of why particular galaxies and stars are and are not targeted. SDSS is a successful existing example of such a collaboration that will provide a new generation of challenges in its next phase.

The data releases of the SDSS projects have been a cornerstone of our success, both in science and public outreach. We will continue our solid tradition of open science and collaboration with other astrophysists, computer scientists, and educators, and enhance our tools for engaging all of them in cutting-edge scientific exploration with large data sets.