The Sloan Digital Sky Survey
Archive Distribution Plan to the Astronomy Community
Revision 2 - April 5, 2001

1. Background

The NSF Astronomy Division provided funds for the construction and commissioning of the SDSS with the expectation that the survey data would be made available to the astronomy community. For that reason, the 1994 Cooperative Agreement AST-9405243 between ARC and the NSF stipulated that ARC would provide an acceptable plan for the distribution of the data to the community. In the fall of 1998 the SDSS management prepared a public data distribution plan and submitted it to the Program Manager for Advanced Technologies and Instrumentation. Following a review, under the Program Manager's direction, the plan was revised and then accepted in April 1999. This plan, the 1999 Plan, continues to be the foundation of our efforts to release the data to the community. It was revised in September 2000 to include the release of the science-quality commissioning data in July 2001, more than a year in advance of the first release that was specified in the 1999 Plan. The release of the commissioning data is called the Early Data Release in order to distinguish it from subsequent releases. The data contained in the Early Data Release do not consistently meet survey requirements and the precision of the calibration will fall short of survey requirements at the time it is released. The calibration of this data should be substantially improved by the time of the first scheduled release of survey quality data in January 2003. Nevertheless, the quality of the data contained in the Early Data Release is excellent and will be of great value to the community. The purpose of this document is to recognize the role of the Space Telescope Science Institute (STScI) in the Early Data Release and the expectation on the part of ARC that the STScI will become the long-term steward of the SDSS Archive.

In November of 2000 STScI agreed, in principle, to collaborate with ARC in the distribution of the SDSS data to the community. The expectations for the roles of the SDSS and the STScI in the distribution of the SDSS Archive can be summarized as follows. The SDSS Data Processing and Distribution team will prepare the data products in a format that is compatible for distribution through the STScI Multi-Mission Archive (MAST). This team will build a set of data servers that will contain the data products. The STScI will then make the data products available to the community through MAST. In order to enable the distribution, STScI will adapt the MAST web interface for use with these data servers. During the observing and data processing phases of the Survey, SDSS will provide technical support for the maintenance and calibration of the Archive and the data servers. After the final release of the SDSS data in July 2006, STScI will maintain the Archive and the data servers. The remainder of this document describes the data products, the schedule, and the roles of the SDSS and the STScI in the data distribution to the community.

The SDSS data products vary in complexity and size. The full volume of data is beyond the capability of most users to store or utilize effectively. Nevertheless, some research projects do not require all of the data from the complete SDSS area and can be accomplished before the observing phase is complete by using subsets of the data that are processed and calibrated as the survey progresses. As another example, a single user can

undertake a significant research project with just the positions and redshifts of "only" one million galaxies. Some projects may require significantly more information about each object than the simple redshift catalog provides, such as the measured photometry of objects and corrected image frames. Some projects may choose to use different calibration procedures or even different processing algorithms. These latter projects require the type of computing facilities that only major computing centers possess. The data distribution plan takes these various possibilities into consideration.

There is also a trade-off between the prompt availability of the data to the community during the survey and the integrity of the calibrations. It is in everyone's best interests to ensure that that all released data have high quality, reliable and, most of all, well-characterized calibrations. This plan attempts to balance these competing requirements by including an Early Data Release.

2. Natural Time-scales – Points of No Return

The interplay between the photometric survey and the spectroscopic survey defines the SDSS observing strategy by imposing two well-defined "points of no return" on the data processing. The first occurs when a portion of the imaging data is determined to be good enough to allow target selection. The second occurs when the spectroscopic reductions are good enough to complete a particular "tile" on the sky. The first event is a particularly hard boundary. It would be very costly to re-image, re-process, re-select, re-tile, re-drill, and re-observe the spectra, if we discovered that the selection of objects was erroneous, after plates have been drilled for specific targets and their spectra obtained. The sky area that the SDSS can cover is limited by funding and if a major portion of the survey area has to be redone it would compromise the entire survey. It is essential for the success of the survey that we have a clearly defined "point of no return" for target selection. In effect, the scientific quality requirements related to the homogeneity of the spectroscopic survey define the timing and other procedures for acceptance of the photometric data. As noted earlier, some of the data contained in the Early Release Data do not meet survey requirements and the entire process for the sky area from which this portion of the data was obtained will have to be repeated. The schedule for the data distribution is referenced to these two points of no return.

3. *Quantized Data Release*

In order to provide a uniform, statistically meaningful data archive, we will release the data in yearly quanta. The complexity of the SDSS data and the need for time consuming, repeated verifications of the calibrations create the latency: the time interval between the time a quantum of data has been first processed and calibrated and the time the processed and calibrated data have been determined to have met survey requirements. This latency also includes the time it takes to prepare the data for export to the STScI and subsequent distribution. When the 1999 Plan was approved, we expected a latency of eighteen months for the first release. This is consistent with our experience and we plan to gradually decrease the latency to less than a year by the fifth year of the survey.

4. *Data Products, Servers and Interfaces*

4.1 *Data Products*

The SDSS data products that will be released to the community, their approximate sizes and the main access mechanism are described in this section and the list of these products is shown in Table 1. The method of accessing the data is noted in the table by the following symbols: **W** denotes web access; **C** denotes the Catalog Archive Server; **D** denotes the Data Archive Server; and **S** denotes the skyServer. Detailed descriptions of the servers and their locations are presented later in section 4.2.

Table 1. Data Products

Product	Size	Access
1. Survey Description and Status	1 GB	**W**
2. Full Object Catalog	1 TB	**C, S**
3. Atlas Images	3 TB	**D**
4. 1-D Spectra	*	**D**
5. Compressed Sky Map	300 GB	**D**
6. Masks	*	**D**
7. Calibrations	5 GB	**D**
8. Corrected Imaging Frames	15 TB	**D**
9. Augmented Color Images	150 GB	**D, S**

1. *Survey Description and Status*

The description of the survey is a set of documents that will reside on the SDSS website. They will be updated automatically, approximately on a weekly basis. Users may access this information through the MAST interface, which will contain the full set of documents and will obtain the dynamically created survey operations status information from the SDSS website. The documentation will include the various file formats used. The status information will include a list of stripes/strips observed to date, the status of photometric data processing, the status of target selection, the status of spectroscopic data processing, the weather logs, and instrument logs. A prototype of the reporting system is already on-line and undergoing tests.

2. *Full Object Catalog*

This product consists of the output quantities from the imaging pipeline (objc parameters) and the 1D spectroscopic pipeline. It includes positions, several different types of magnitudes, shape parameters and their errors for all objects in 5 bands within the detection limit of the imaging survey. It also provides radial profiles in logarithmic bins and their errors, survey coordinates and pixel coordinates. The

spectroscopic data in the catalog contains quantities such as redshift, classification and line strengths for all targeted objects. The targeted objects include galaxies, quasars, stars with various properties, and sources from the ROSAT and FIRST catalogs. The Full Object Catalog also contains all relevant calibration information for the objects. It can be accessed through the Catalog Archive Server. Astronomers will be able to download selected data to their site and make their own redshift and image catalogs from this server.

3. Atlas Images

Atlas Images are cutouts of images of detected objects extracted from the full image frames in 5 bands. They amount to over 1 billion images. They will be available via a web interface, and will be accessed either by their locations or by their unique SDSS identification. Individual images will be available on-line and can be linked via their own URL. Images can be accessed either as single FITS images, an array of FITS images, or as augmented color thumbnail images for browser preview. This product will be accessible from the Data Archive Server.

4. 1-D Spectra

The 1-D Spectra are the calibrated spectra of each spectroscopically observed object, together with a noise array and a mask array. They will be obtained as a FITS file through the Data Archive Server. Spectra can be retrieved by a simple web-based query interface, or by their sky position or their SDSS identifications.

5. Compressed Sky Map

This product is a 4x4-compressed version of the set of all image frames after detected objects have been removed. This map, along with the atlas images can be used to reconstruct a good approximation of the original image frames. This will be provided through the Data Archive Server soon after the Early Data Release.

6. Masks

The Masks include the set of flags set for each pixel during the photometric processing that define in detail the parts of the image frame that could or could not be processed for various reasons. They also include the set of flags from the 2D output of the spectroscopic pipeline. The Data Archive Server will provide them.

7. Calibrations

This product contains the astrometric (position) and photometric (flux) calibration coefficients for the full image frames along with miscellaneous other information, such as the seeing and the sky background. These will be versioned, so that the actual calibrations used for target selection will still be available even after additional work produces better calibrations. The calibration parameters, including their versions, and various other instrumental records will be served as FITS files from the Data Archive Server after the Early Release.

8. Corrected Imaging Frames

The corrected, flat fielded imaging frames which were used for object detection, with updated WCS information will be provided as FITS files, and can be accessed from

the Data Archive Server, with a simple query interface. They can also be directly linked via their unique URL.

9. ***Augmented Color Images***

A three-color image for each frame in the survey will be generated for easier previewing with a web-browser. These Augmented Color Images will be available in at least four zoom levels and will be stored in JPEG format. Users will be able to make queries to select these images and their URLs will be available for direct linking applications. A color GIF image of the 1-D spectrum for each spectroscopically observed object will also be served by the Data Archive Server. The images will show the flux as a function of wavelength, with the detected lines and features marked in color.

4.2 *Servers*

Three platforms, the Data Archive Server, the Catalog Archive Server and the skyServer, will be used to distribute the different data products to the community. All SDSS data will be accessible either via web browsers, or a simple, downloadable lightweight Java application.

The Catalog Archive Server

This server employs the SX software (Science eXplorer), which was developed by JHU, built on top of a commercial object-oriented database, Objectivity/DB. The Catalog Archive Server contains the full object catalog and it enables the user to carry out complex queries over all the output quantities of the photometric pipeline (objc parameters) and spectroscopic pipeline (spectro 1d output parameters), plus the calibration parameters used in the processing. SX enables queries through a Java based query tool, which uses a query language close to the standard SQL with some astronomy specific extensions that were developed by JHU. In the future, an Applications Programming Interface will also be provided to enable users to retrieve data directly from their data analysis packages. This will be coordinated with the NVO effort, to make sure that it will comply with the emerging new standards.

The Data Archive Server

This server uses a simplified version of the SX server, developed by JHU, that provides access to approximately 3 Terabytes of atlas images, spectra, corrected frames, compressed sky map (binned images) and the masks. The Data Archive Server can carry out simple queries, based on right ascension and declination or object identifications. It will retrieve the processed data in FITS format, through a web-based interface adapted from MAST. The data can be downloaded either though the query form or through a direct URL pointing to the respective data files. It also contains detailed supporting documentation about both the data products and the SDSS project itself.

The skyServer

This server has an interactive web-based interface, which provides an integrated navigation of images and the Full Object Catalog. The skyServer is built in the Windows environment, using Microsoft's SQL Server database. It contains the Full Object Catalog, with redshifts and spectral lines, a full set of color images in compressed JPEG format and GIF images of all the spectra, with detected spectral lines graphically marked. This server was developed jointly by Microsoft's Bay Area Research Center (BARC) and JHU. It is targeted to support the K-12 audience and the general public, but it has some features that astronomers might also find useful. It will be used as our primary mechanism to provide content for educational and outreach efforts using SDSS data.

4.3 *Interfaces*

Astronomers may obtain access to the SDSS data products through three interfaces; MAST web interface, sdssQT, and skyServer. The expectation is that the MAST interface will be the preferred interface. The relationships among the different views of the SDSS data can be seen in Figure 1. The color-coding in this figure shows the institutional responsibilities for the development of the interfaces for the Early Data Release. The physical locations of the servers at the time of the Early Data Release are shown later in Table 2. Section 4.4 describes the institutional responsibilities for the maintenance of the servers.

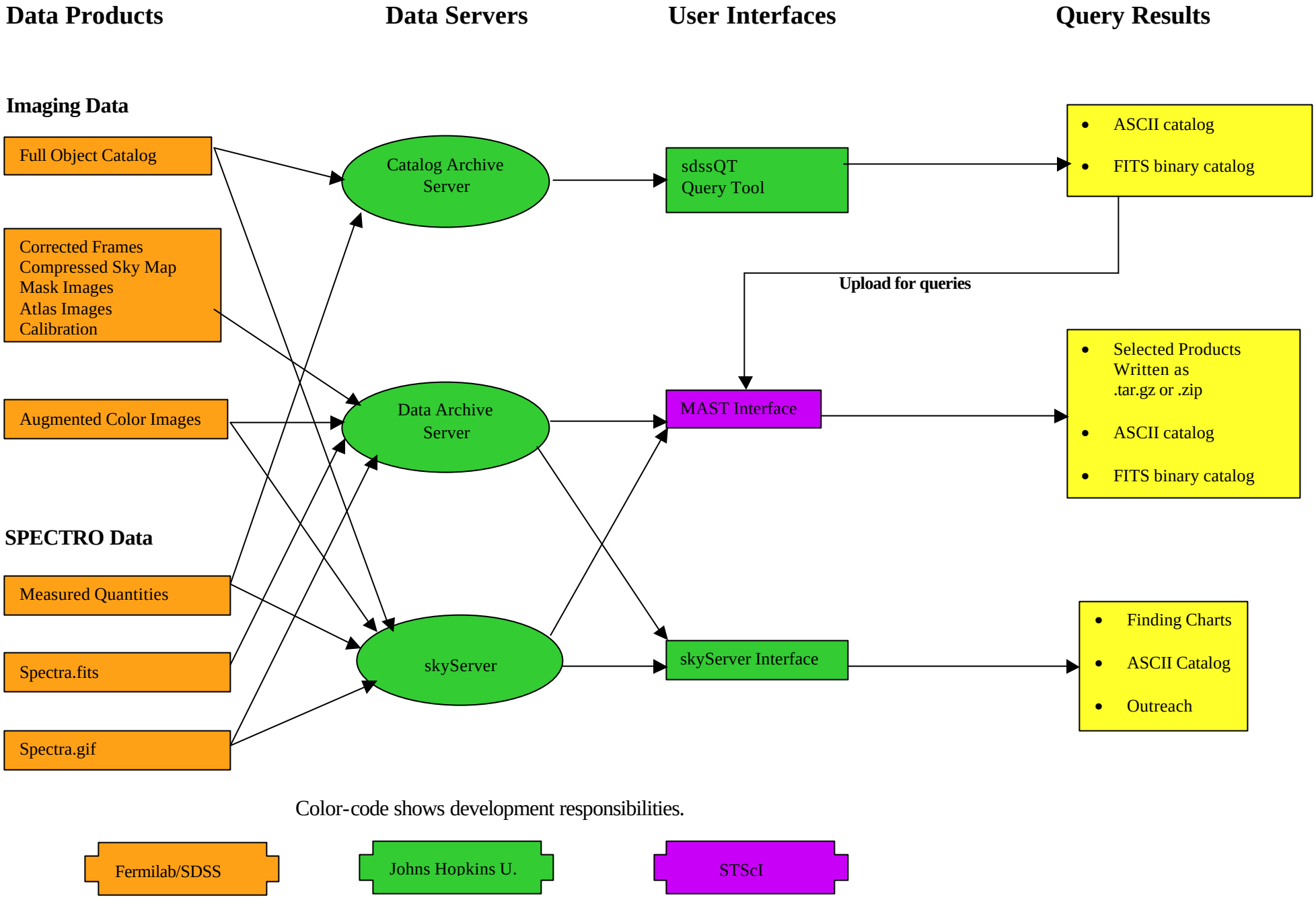

Figure 1. Data Flow Chart

STScI has adapted its MAST interface for use with the servers and data products developed by the SDSS. It has also developed detailed documentation, user guides and help pages. It has distributed the software, documentation and web pages to Fermilab as well. Local copies of the MAST interface will run at Fermilab, but the primary access to the SDSS data will be through the MAST links. The skyServer interface was designed and is being developed jointly by Microsoft BARC and JHU.

4.4 *Location of Servers*

At the time of the Early Data Release, web accessible data servers will be located at Fermilab and the STScI. Since it is anticipated that the servers at these sites will not efficiently serve data to astronomers in Japan, the JPG is building a mirror site to serve astronomers in Japan. STScI will make its interface available to the JPG, when it is ready to accept it. The JPG will be responsible for any custom modifications. The Max-Planck-Institutes participating in the SDSS are considering creating a mirror site to serve astronomers in Germany, although it may not be ready by June 2001. Table 2 summarizes the types of servers and their locations at the time of the early data releases

Table 2. Locations of Servers in July 2001

	Data Server	**Catalog Server**	**skyServer**
STScI/MAST	Yes	Later	No
Fermilab	Yes	Yes	Yes
Japan	Yes	Yes	Yes

1. At the time of the Early Data Release, STScI will host a Data Archive Server, which will contain the atlas images, the corrected and reconstructed frames and all outputs of the spectroscopic pipeline. As other products such as the calibration and masks become available, they will be added. The STScI will support its Data Server with existing resources. The Data Archive Server hardware is already in house, and is being tested. The STScI plans to install a Catalog Archive Server after the Early Data Release.
2. Fermilab will host a Data Archive Server, a Catalog Archive Server and skyServer. These servers are installed and being testing. Fermilab acquired the hardware and the appropriate software licenses for the Data Archive and Catalog Archive Servers with DOE funds. Microsoft BARC and Compaq Computer Co. provided the necessary resources for the skyServer.
3. Japan will set up and maintain its own copies of the Data Archive Server, Catalog Archive Server and skyServer. The latter will be translated to Japanese for Japanese viewers. The JPG will acquire and support the servers in Japan with independently obtained Japanese funding. The JPG has placed their orders for the necessary hardware, which will be configured as an exact replica of the systems at Fermilab.

4. There may also be a mirror site in Germany. The funding and support of these will be the responsibility of the Max Planck Groups in Munich and Heidelberg. The details for these systems are not yet available.

4.5 *Support and Maintenance*

All interfaces and services will be under strict version control in order to keep the interfaces and the servers synchronized with the master copies. The STScI will maintain a Help Desk, which should be able to answer most of the issues raised by users. It will assign those questions and requests it cannot handle to the SDSS Data Processing and Distribution Team for action. This group will use the GNATS problem report database to assign responsibility for action and this system is already in place.

4.6 *Development and Testing*

Each type of server will go through several stages of development. The first, the prototype stage, contains incomplete data sets that were processed with earlier versions of the pipelines and has preliminary calibrations. The purpose of the prototype stage is to evaluate the performance of the interfaces and the server software. The SDSS Data Processing and Distribution team is responsible for this evaluation and astronomers from the SDSS Collaboration are responsible for the evaluation of data contained in these servers. STScI will be responsible for creating and testing the MAST interfaces.

After the operational problems have been fixed, a beta version of the servers and their data will be released to the SDSS Collaboration. When this milestone has been reached, the servers will be functional, and the data will have been processed with versions of the pipelines that will be used for the Early Data Release. At this point the data will have been calibrated with the final calibrations for the Early Data Release. The major purpose of the beta release is to validate the quality of the data connected to each server. In the event that problems are found with the beta data set, they will either be documented on an errata web page, or if possible fixed. In particular, it will be possible to update the calibrations before the production version is released. The SDSS testing team will be responsible for data validation. STScI will be responsible for making any necessary improvements to the MAST interface.

After the data has been validated and any serious, recently discovered software bugs have been fixed, the production version will be released to the community. Throughout these development and testing phases, the servers at all sites will be available to the SDSS Collaboration. The intent is to expose as many operational problems as possible before releasing the data and the servers to the community.

5. The Schedule for the Release of the Data to the Astronomy Community

The schedule for releasing the data to the community is shown in Table 3. The dates of Releases 1 through 4 are identical to the dates contained in the 1999 Plan. While regular accumulation of photometric data began in April 2000, three months later than we forecasted when the 1999 Plan was prepared, we believe that these dates are achievable.

Table 3. Dates for the SDSS Data Releases

	Release	Photometry	Spectroscopy
Early Release	1-July-2001	5%	0%
Release 1	1-Jan-2003	15%	7%
Release 2	1-Jan-2004	47%	33%
Release 3	1-Oct-2004	68%	60%
Release 4	1-July-2005	88%	85%
Final Release	1-July-2006	100%	100%

Regular accumulation of photometric data will end in the summer 2004, except for some limited opportunities during the last six months of the survey. As the observing phase draws to a close, the opportunity to take imaging data, carry out target selection, drill plates and take spectra of the associated portion of the sky rapidly diminishes. Spectroscopic observations with fully commissioned spectrographs began in April 2000 and will continue to the end of March 2005.

Figure 2 shows the planned rate of accumulation of the two main data components, the photometric data and the spectroscopic data. The intermediate milestones occur on July 1, in each year starting in 2001. Two other critical milestones are the beginning of the survey, and the end of survey observations. The planning assumptions that determined these milestones are: observations of the northern sky can be made during the first two quarters of every year, the third quarter is largely lost to the monsoon season, and the northern galactic cap can be observed for only about one month during the fourth quarter. The data obtained prior to the "points-of-no-return" is quantized by the mid-year milestones and will be released at the times shown by the tip of the arrows. Somewhat arbitrarily the date of January 1, 2000 was shown as the effective starting date of the survey in Figure 2, because it simplifies the graphical presentation. It also reflects our expectation that some of the commissioning data will meet the Science Requirements after it is properly calibrated.

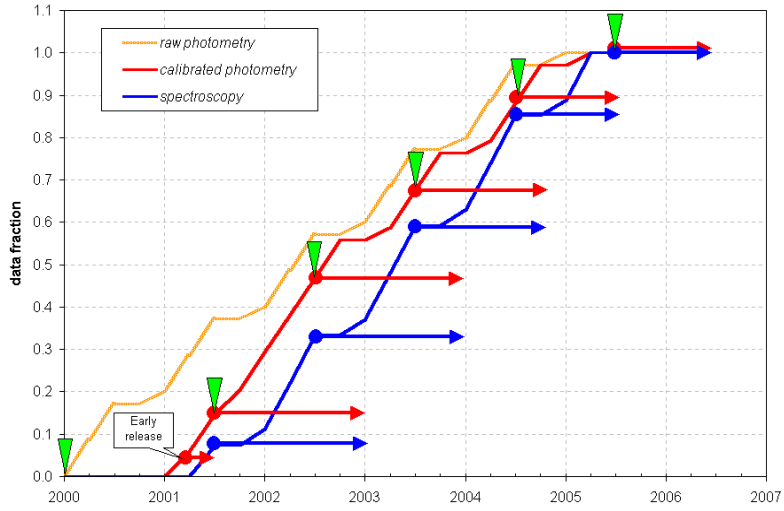

Figure 2. Milestones and data fractions for the new SDSS Distribution Plan

The Early Data Release will contain over 400 square degrees of imaging data on the equator, in both the Northern and Southern skies. It will also contain about 60 square degrees in the Northern Galactic Cap that were observed in Spring 2000. A portion of these 60 square degrees has already been provided to the SIRTF program for the First Look Survey (FLS). The Early Data Release will also contain over 50,000 spectra of objects that were selected from the imaging data that will be contained in the Early Data Release.

6. *Organization and Status of the effort for the Early Data Release*

The tasks for the Early Data Release were assigned to STScI, JHU, and Fermilab according to the resources and experience that each team could bring to bear on short notice. The primary goal of this effort is the distribution of the high quality commissioning data to the community in July 2001 in a format that is familiar to many astronomers. The Early Data Release will be accomplished using the existing resources available to the SDSS and the STScI. NASA has allowed STScI to support the Early Data Release within the existing resources of MAST. STScI management does not plan on expending more than 0.75 FTE year on the Early Data Release. The goals for distribution of the SDSS Archive to the community after the Early Data Release has been completed are described in section 7.

Since astronomers are familiar with MAST, the task of adapting its user interfaces to the data servers was assigned to the STScI/MAST group. They have made excellent progress toward prototyping this interface to the SDSS Data Products within the SX environment.

The MAST team is adapting their existing feedback mechanism to link the astronomer-user to the SDSS technical staff when problems arise. It begins with the STScI call tracking system and it allows the MAST staff to receive requests and route them to the appropriate organization for action. Items that cannot be handled by the STScI will be sent to the head of the SDSS Data Processing Distribution team who will enter the problem in the SDSS GNATS problem report database. The GNATS problem reporting system has been in place for many years and SDSS astronomers are familiar with it. It will be tested in this new environment in the very near future. The exploitation of existing resources has allowed the creation of the user interface to go forward much faster than if the SDSS had to do the work by itself.

The job of adapting the SX software and the SDSS data model to the MAST framework was given to the JHU development team, the only members of the SDSS-STScI collaboration with extensive experience with SX. The proximity of the JHU and STScI groups has allowed this work to progress very rapidly. The MAST group could not undertake this task without additional resources, given the requirement to complete the work in a very short time and their need to undertake an extensive learning effort to come up to speed on the SX servers. The assignment of this task to JHU was the right choice given the need for changes to the data model that required changes to the basic SX framework. There will be further changes to the data model after the Early Data Release

in order to incorporate the calibration data into the Data Archive Server and to accommodate planned improvements in the imaging pipeline.

The SDSS team was assigned the responsibility for the creation and collection of the documentation. They are making extensive use of documentation that is already on the SDSS web site. The STScI is making rapid progress on the framework for distributing the documentation of the SDSS Archive through web pages. The SDSS team is already using MAST web pages to set up user interfaces at Fermilab and JHU. The MAST group has already linked MAST to the existing SDSS web pages.

The job of preparing the processed data for export to the servers at each institution was assigned to Fermilab since they have the most experience with this step and it fits naturally with work they are already doing. The Fermilab data processing and distribution team is also responsible for commissioning the servers at Fermilab and loading the SX databases at Fermilab with the appropriate data sets. This is similar to the work that they have already done to load the chunk database and the rerun databases, which are similar in design to the Catalog Archive Server. They are now operational and are being used by the SDSS Collaboration to validate the data. Fermilab will also be responsible for partitioning the Data Archive Server database into a public section and Collaboration section. The Collaboration section will contain data that is not ready for distribution to the community.

The task of developing the WBS and schedule for the Early Data Release was assigned to the SDSS Project Manager and his Office. The Project Manager's Office updates the task list weekly and reports the critical path.

The SDSS pipeline developers have been and will continue to be responsible for developing and evaluating the Quality Control tools. They are giving special attention to the Early Data Release, since the 2.5-m light baffles were not in the design configuration in 1998 and early 1999. A small, special pipeline is being built, commissioned and tested by the Princeton software development group in order to accommodate this data. After first level quality assurance tests are completed at Fermilab and Princeton, astronomers throughout the Collaboration will use the Data Archive Server and Catalog Archive Server to test and validate the Early Data Release products. The Early Data Release has been given the highest priority within the sphere of testing and validating activities. SDSS astronomers from Chicago, Fermilab, JHU, Princeton, and the University of Washington have already begun this work and their efforts will also serve to product-test the data servers.

7. After the Early Data Release

The SDSS team is undertaking its contribution to the Early Data Release with the existing work forces from the Participating Institutions by deferring some work on data taken in late 2000 and 2001. Since this data will be released as part of Release 1 in January 2003, the amount of data released on 1 January 2003 may be less than forecast in Table 3. It is important to note that Survey Operations, including the creation of the data archive and its subsequent distribution to the community, is not fully funded. ARC has requested funds from the NSF for this task through pending NSF proposal number

0096900. The creation of the SDSS-STScI collaboration made it possible to transfer the responsibility for the distribution of the Archive to the community from the SDSS team to the STScI. The STScI has provided NASA with an estimate of the cost and the required resources for this task and has requested the funds from NASA to continue the work after the Early Data Release. Although NASA endorses the distribution of the SDSS data via the STScI, NASA has not yet awarded funds to permit the STSCI to do further work.

The SDSS Team must continue to carry out all of the tasks up to the final distribution of the Archive to the community after the Early Data Release. The transfer of these tasks to the STScI would not be cost effective while the SDSS Team is acquiring and processing data.. The SDSS Team will continue to be responsible for all aspects of creating the SDSS Archive. In particular, it will continue to carry out Quality Assurance tests on newly acquired data before it is exported to the STScI for distribution. If correctable problems are found in already exported data before the final data release, the SDSS will reprocess the data and then export the reprocessed data to the STScI. The STScI is expected to take over this responsibility after the final release. Prior to that time it will estimate the necessary resources and request funds for these purposes.

The SDSS will continue to be responsible for generating reports of the status of observations, data processing, and data export. The SDSS team generates the information for these status reports in the course of doing its work. It will be most efficient to distribute these reports to the STScI through the SDSS website. In addition the SDSS will continue to use its website for the distribution of data and other technical information to the SDSS Collaboration and for outreach to the general public by ARC. In order to carry out these functions more effectively, the SDSS team will need to increase the level of effort that is committed to the completion of its website. It will not be cost effective to assign these jobs to STScI, although it may be possible to adapt some of their software tools to this effort, thereby keeping the additional costs to a minimum.

The SDSS will provide the STScI with improved calibrations of data that has already been exported to the STScI and will provide additional data to the STScI in accordance with Table 2. As noted earlier, the former task will be done by reprocessing previously exported data with new calibrations and then re-exporting the data to the STScI. After the Final Release, it is intended that this responsibility will be transferred to the STScI, since the SDSS Collaboration may have been disbanded. The SDSS and STScI will take the appropriate steps to make it possible for the Archive to be calibrated independently at some time in the future. The Fermilab EAG group and the Princeton software group will be responsible for these SDSS tasks during the observing and data processing phases of the survey. The cost and level of effort required to carry out this work is described in Section 2 of the revised Management Plan provided separately to the NSF.

The JHU SX development group will continue to support and improve the Data Archive and Catalog Archive Servers. These systems have not been tested under a heavy load, and it is expected that experience with the Early Data Release will make it possible to make an assessment of the needs for subsequent releases. As the amount of data increases, multiple servers (nodes) will be brought on line. While parallel operation of nodes is part of the design, it has not yet been necessary to utilize these features. The current operating system for the servers, Linux, is undergoing rapid evolution and the

SDSS team is prepared to make changes to keep pace with system changes. Based on our past experience, just maintaining the current software (over 100,000 lines of our own code) will require 1 FTE until the end of the Survey. Furthermore, it is expected that users will want enhanced capabilities to extract their results as they gain experience with large data sets. This effort will require another FTE until near the end of the survey. Further details of the required resources that the JHU team will need are contained in Section 2 of the revised Management Plan.

The STScI is willing and able to become the long-term steward of the SDSS Archive once it has been assembled. After the Early Data Release, the existing hardware at the STScI, particularly disk storage, will not be adequate to manage the entire SDSS Archive and the other archive responsibilities assigned to STScI by NASA. The Early Data Release will represent slightly more that 5% of the total SDSS Archive. The STScI has requested funds from NASA to acquire the hardware and software, including licenses, to host the full SDSS Archive and to manage it as it grows and the STScI responsibilities grow with it. After support from the SDSS is no longer available, the STScI plans to acquire the expertise to maintain and improve the SX servers developed by the SDSS should they prove to be an effective way to distribute the Archive. The STScI intends to assume full responsibility for distributing the SDSS Archive to the community after the January 2004 data release. By the time of that release, the STScI will be distributing about 50% of the final Archive to the entire US astronomy community. While Fermilab will still be able to provide some service to the community, it will size its hardware and software to distribute the SDSS Archive to the SDSS Collaboration, which is considerably smaller than the US astronomical community.