Fermi National Accelerator Laboratory

FERMILAB-Conf-95/064

# The Sloan Digital Sky Survey's Use of the Web

D. Petravick, E. Berman, V. Gurbani, S. Kent, T. Nicinski, R. Pordes, R. Rechenmacher and G. Sergey

*Fermi National Accelerator Laboratory*
*P.O. Box 500, Batavia, Illinois 60510*

R. Lupton and M. Richmond

*Princeton University Observatory*
*Princeton, New Jersey 08544*

April 1995

Presented at *Weaving the Astronomy Web Conference*,
Strasbourg Astronomical Observatory, Strasbourg, France, April 6-7, 1995

## Disclaimer

# The Sloan Digital Sky Survey's Use of the Web*

Don Petravick, Eileen Berman, Vijay Gurbani, Steve Kent,
Tom Nicinski, Ruth Pordes, Ron Rechenmacher, Gary Sergey

*Computing Division, Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, U.S.A.*

Robert Lupton, Michael Richmond

*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, U.S.A.*

Fermilab, as one of the Sloan Digital Sky Survey collaborators, primarily uses the World Wide Web to disseminate information about project software and software engineering practices.

## 1 Introduction

Fermilab, as part of its Experimental Astrophysics program, is collaborating in the Sloan Digital Sky Survey (SDSS) [1]. Fermilab has major responsibilities for the project's software and software engineering practices. The dissemination of information about these to the SDSS collaborators is primarily done through the World Wide Web.

The survey will collect and reduce 12 TBytes ($12 \times 10^{12}$) of data. The data will be collected over a period of 5 years at Apache Point Observatory in New Mexico, starting in late 1995. The imaging data, taken in 5 colors, will come from a photometric survey of $\frac{1}{4}$ of the sky. Over 10 million galaxies and stars each will be imaged. Based on the photometric data, the survey will concurrently collect 1 million spectra of galaxies and 100000 spectra of QSOs. The resulting data sets are expected to provide interesting science for 20 years.

* Sponsored by DOE contract DE-AC02-76CH03000 and the Alfred P. Sloan Foundation.

## 2 SDSS and the World Wide Web

The SDSS is a world-wide collaboration with contributors from around the United States, Italy, and Japan. Collaborators need to have easy access to current information about the project and the on-going software development. The World Wide Web (WWW) [2] and HTML (HyperText Markup Language [3]) meet this need admirably.

Early in 1993, the collaboration established the World Wide Web as *the* method of information dissemination. The Survey was one of the early users of WWW. It has relied on outside software for browsers and infrastructure, like server frameworks. Much of the Survey documentation is written in HTML, rather than being converted from another format. Because it was an early user of HTML, and because HTML is the primary language for much of our documentation, we developed, and continue to use several custom tools.

Our use of the Web can be thought of as providing access to three distinct types of information: 1) the infrastructure of the project, 2) the documentation of actual software used throughout the project, and 3) access to Survey databases. The layout of Web pages is not meant to allow the public to easily surf. Rather, the pages are laid out to permit collaborators to find information efficiently. In keeping with this goal, there is less emphasis on glitz and more on content.

## 3 SDSS Infrastructure

Although each collaboration site manages its own Web pages, with Fermilab's major role as a software provider (both data acquisition and infrastructure), Fermilab maintains a central WWW server providing access to much of the SDSS' online documentation. The WWW server is also used to bind project documentation from all sites.

### *3.1 Introductory Information*

As part of the SDSS infrastructure documentation, we maintain an introductory section that is directed at people new to the project, providing such information as the goals/overview of the project, and tutorial sections to help one become familiar with the SDSS software.

### *3.2 Instrumentation*

Information about what's new in the Survey is posted here. The survey is currently gathering test data and commissioning its instruments. Scientists at the observatories post news on the Web, where it is available to the collaboration.

We also use the Web to provide services to our observers. We have the Guide Star Catalog (GSC) available on the Web. It may be queried using an HTML form to produce finding charts.

### *3.3 Who's Who*

Through our central server we also maintain several lists. Names, mail addresses and phone numbers of all the collaborators and their associated institutions are accessible through the server. Additionally, SDSS network cluster information and mail alias lists are available.

### *3.4 Top Level Indices*

Although the SDSS Web pages are hierarchically oriented, some users prefer a flat listing of available topics for quicker maneuvering through the Web. For this reason, an index of top-level documentation is automatically generated from selected HTML documents.

### *3.5 Protection*

Of course, certain documentation is sensitive in nature and is made unavailable to the Internet public via standard WWW server access protection mechanisms. Although we utilize these protections, we maintain the attitude that diligent hackers may still be able to penetrate, and as such, any extremely sensitive material should not be placed on the Web.

## 4 SDSS Software and Documentation

The SDSS has over 1 million lines of software, which is developed by many of its collaborators at different institutions. Use of the Web allows all of our

collaborators to view our current documentation without taking delivery of the actual software.

SDSS software is not monolithic, but rather consists of many "products," each with its own independent set of source files and documentation. Since products may have developers from several collaboration sites, the RCVS (*R*emote extension of *C*oncurrent *V*ersions *S*ystem) code management tool [4] is utilized to coordinate modifications.

RCVS permits people around the world to modify code, and its associated documentation, locally. Users "check out" a copy of a product to a local work area. They modify the code and documentation, compile it, debug it, and then "commit" those changes back into the RCVS repository. That makes those changes available to the rest of the collaboration.

The use of RCVS permits collaboration contributors to easily add material. They can also modify and correct documentation, promoting improved dissemination of accurate information because of the Web's ability to provide immediate access to the latest documentation.

### *4.1 Documentation Within Source Code*

Initially, all documentation was written directly in HTML. Recently however, there has been an effort to incorporate HTML directly into the software source code, thus avoiding duplicity of documentation and reducing maintenance. It was our goal to adequately document software routines and their parameters/prototypes. Since much of our coding conformed to a particular style of commenting, our efforts to incorporate HTML were hastened by creating a utility to automatically extract appropriate sections of code (such as routine prototypes) and build HTML.

### *4.2 URLs and Mappings*

URLs fixed to hardwired paths based on a product version is a concern. Since our coding standards restrict a product's HTML documentation to a single directory, relative URLs are used. For cross-product references, server configuration file mappings are available for each SDSS software product. This allows URLs to contain only the product name, void of any version specifics. As new product versions are developed, only the mappings in the configuration file need be changed. Naturally this means only the latest set of documentation is visible through our servers, but fortunately this is sufficient for our needs. Older versions can still be viewed by opening files locally with a browser.

Currently the server configuration file is managed by hand, but plans are in the works to automate the process, so whenever a new version of our software is produced, the configuration file is updated and reread automatically.

### *4.3 Product Indices and Search Capabilities*

As indicated, our primary goal concerning documentation of our software is to produce descriptions of all user callable subroutines defined in the product libraries. As such, indices and search capabilities are a useful part of our documentation. For our needs, a general text search capability was not required. We found it adequate to simply search among all defined HTML anchors (<A NAME ...>) and title tags (<TITLE ...>). Thus, a simple script was developed to perform both the searches and to generate indices.

### *4.4 Software Compilation Reports*

Since our software is destined to be run on a wide variety of UNIX-based platforms, our quality control mechanism routinely compiles our software on all supported systems. Compilation messages are filtered such that messages needing attention pass through and are incorporated into an HTML report. This hypertext report allows quality control personnel to view lists of messages sorted by operating system or source file by following appropriate links. In addition, the modification history of any of the offending source can be accessed since the script also interacts with the RCVS utility.

### *4.5 Bug Reports*

The Web is used as a means to track product bug reports and their status. The publicly visible bug list lets people know about problems as they are discovered. Rather than burden users with details, the list is a summary of the problems; details can be gotten by following links.

### *4.6 Documentation for the Non-Web World*

Finally, an HTML to LaTeX converter is used to generate printable documentation.

## 5 Access to Survey Databases

SDSS currently uses the VERSANT™ object-oriented distributed database [6] to maintain various catalogs of data used throughout the collaboration. Convenient access to these is provided through a set of hierarchical HTML forms: selecting the database, selecting a class within the database, providing selection criteria to locate class instances, etcetera. The database requires accesses to it to be made in the scope of a session from the same task. As the server is stateless (it does not maintain state information between form submissions), initiating a session for each query is too time consuming. Rather, the initial form spawns a persistent task which is used to maintain the context of the database session.

Databases are not restricted to those generated by the Survey. As an example, the Guide Star Catalog (GSC) [5] is available. Providing the coordinates of a star $(\alpha, \delta)$, an epoch, a field size, and a magnitude limit (Figure 1a), one can get an inline finding chart of that portion of the sky (Figure 1b).

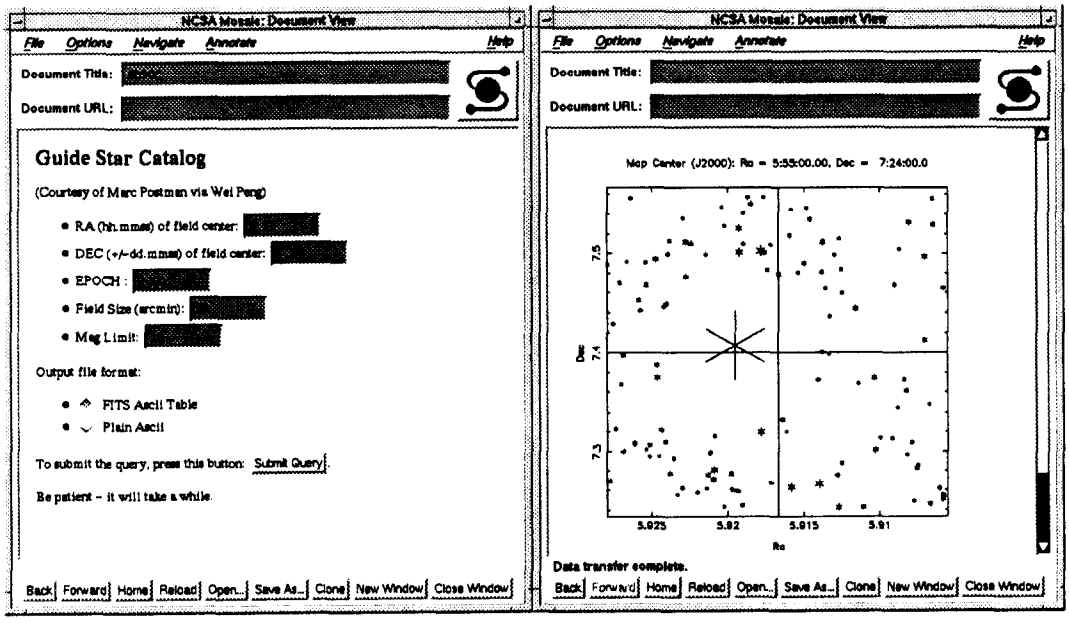


Fig. 1. a) GSC Query near Betelgeuse and b) Resulting Chart

## References


[1] *A Digital Sky Survey of the Northern Galactic Cap* (December 20, 1990).

[2] T. J. Berners-Lee, R. Caillan, J. Groff, and B. Pollermann, World-Wide Web: The Information Universe, *Electronic Networking: Research, Applications and Policy* Vol. **2** No. 1 (Meckler Publishing, Westport, CT, Spring 1992) 52–58.

[3] D. Connolly, *HyperText Markup Language (HTML)* (*http://info.cern.ch/hypertext/WWW/Markup/Markup.html*).

[4] T. Hung, *RCVS: Remote extension of Concurrent Versions System* (Stanford Linear Accelerator Center, unpublished document).

[5] Space Telescope Science Institute, *The Guide Star Catalog* CD-ROM (Association of Universities for Research in Astronomy, Baltimore, MD, 1989).

[6] *VERSANT System Manual* (Versant Object Technology, Menlo Park, CA, 1992).