Paper 126-25

Providing Access to an Academic Data Warehouse Over the Web

Phil Rhodes, Baylor University, Waco, TX
Faron Kincheloe, Baylor University, Waco, TX

ABSTRACT

Providing timely and accurate information is the primary function of the Office of Institutional Research and Testing at Baylor University. With the advent of the Web, putting that information at fingertips of administrators and decision-makers is easier than ever before. Using The SAS System®, a web-based Enterprise Information System was developed to allow University decision-makers to evaluate various programs and academic departments over time. This system serves as the access point into the University data warehouse, and includes graphical and tabular reports with drilldown and reachthrough capability into individual student records. The only requirement on the client side is a web browser, with Java required to view the interactive graphics.

INTRODUCTION

The Office of Institutional Research and Testing (IRT) at Baylor University is charged with providing timely and accurate information for decision support to the University administration. This information ranges from student enrollment and retention to faculty workload and compensation, as well as data for the University's new Five Year Budget Process. A decision was made to build a University-wide data warehouse in order to consolidate this information and to provide a single access point to the data. The responsibility for implementing the project was assigned to IRT, due to the knowledge IRT had of the necessary data. Using The SAS System, a web-based Enterprise Information System (EIS) was developed to allow University decision-makers to access data contained in the data warehouse.

THE DATA WAREHOUSE

The development of the EIS mirrored the development and implementation of the data warehouse, in that the initial phase of the data warehouse provided access to data for the Graduate School of the University. The Graduate School was chosen because of a new emphasis on research across the University. The current focus of the EIS is also on the Graduate School and on allowing the Dean to evaluate graduate programs based on criteria such as enrollment, test scores, graduation rates, and placement rates. These criteria were obtained from the Dean during the initial planning phases of the data warehousing project.

The data warehouse will eventually contain a wide range of data of very different types. This data will include individual student records, with information on courses, standardized test scores, and grade point averages, as well as demographic information and financial aid data. At the time of this writing, the majority of the data in the data warehouse was the student data, which was the easiest to obtain. Human resource data will also be added, including information on faculty publishing and grantsmanship. Finally, departmental financial data will be included to allow the Graduate Dean to track resource use and allocation.

As is common in many organizations, the data was located in several major repositories, as well as many smaller ones. The primary data source used to fill the data warehouse was the archive of census files created and maintained by the Office of Institutional Research and Testing. These files are extracted from the University's Student Information System (SIS) and archived on the twelfth class day of each academic semester in

SAS datasets. They are used for official reporting purposes by the University. Additional archived data from SIS included monthly and yearly files of applicants to the University. By using these archives to fill the data warehouse, consistency with previously published information was made more likely.

The archived census files have several hundred variables, with many of the variables consisting of short alphanumeric codes. Thus, the large format library developed for these files was part of the initial load of the warehouse. Other major data sources were the University's Human Resource and Financial systems. Smaller data sources included spreadsheets maintained by various academic and administrative departments containing information of interest (for example, the Graduate School tracked stipend sources and amounts using an Excel spreadsheet).

Since a large amount of data was already stored in SAS datasets, converting them to another database format would not have been efficient. Instead, they were left in SAS format and the SAS/Warehouse Administrator® was used to manage the initial data extraction and load of the warehouse, especially the large amount of metadata generated. The Administrator was also used to add additional data to the warehouse and to generate the Multi-Dimensional Databases used in the EIS.

To store the large amounts of data required by the project, a departmental server was purchased for IRT. This server uses the Microsoft Windows NT® operating system. After the decision was made to deliver access to the warehouse over the University intranet, this server was also used to host the EIS.

ACCESSING THE DATA WAREHOUSE

REQUIREMENTS

In planning for providing access to the data warehouse, three main features were required:
- Flexibility
- Ease of Use
- Ease of installation and maintenance

The EIS had to be flexible to accommodate a wide variety of users. On the academic side of the University, these users range from administrative assistants in individual academic programs to the academic Deans and the University Provost. On the administrative side, users include research analysts in IRT, budget and financial analysts, and the University Vice Presidents and the President. Thus, the EIS had to be able to accommodate a wide variety of user types, giving each user access to the information they need to do their job.

This wide variety of user types with a wide range of computer skills also mandated that the EIS be easy to use. Since most users were already familiar with graphical user interfaces (in the form of the Microsoft Windows® and the Macintosh® operating systems), providing a point and click environment for most tasks was added to the requirements. Online help and data dictionaries were made available from all sections of the system. Additionally, tutorials were written to familiarize users with the system and with common tasks.

Finally, the system had to be easy to install and maintain. The project began with minimal assistance from the Information

Technology Center (ITC). Without ITC support, it was not possible for the small staff of IRT to install software on the computer of each user. Again, due to staffing constraints, the system had to be easy to maintain and extend. The University computing environment included mostly Macintosh computers on the academic side, along with a mix of Macintosh and Windows computers on the administrative side, and this limited the choice of software. This desire for ease of maintenance led to an additional requirement that little or no software be installed on the client side.

After evaluating several different options, the decision was made to deliver access to the data warehouse over the University intranet, and that the SAS/Intrnet™ product would be used. This package met all three of the requirements listed above, and allowed the Institutional Research staff to make use of existing SAS programs and knowledge.

DEVELOPMENT

Taking advantage of existing SAS programs, a prototype application was developed to demonstrate the capabilities of the system. This initial version of the EIS used the SAS Application Dispatcher component of the SAS/Intrnet package, along with the Web Formatting Tools, to dynamically generate web pages. The prototype provided access to student demographic and enrollment data only.

The initial user request to enter the EIS was sent to the Application Dispatcher, which used a SAS program to dynamically generate an HTML form based on the user's access privileges. This form reflected the actual contents of the datasets. After the user chose the academic programs and statistics to display, submitting the form requested the Application Dispatcher to run another program to dynamically generate the required report and return it to the user's web browser using the Web Formatting Tools.

This initial interface was demonstrated for the Graduate Dean and, while possibly adequate for some users, was not sufficient to meet the needs of the Dean and his staff. They had additional requirements:

- Drilldown capability, through both the time and organization dimensions, all the way into the detail data
- Dynamic reports that could be reformatted and displayed with different statistics and/or variables
- Export capability into Microsoft Excel for additional analysis

It would have been difficult to meet these requirements using only the Web Formatting Tools and the Application Dispatcher, due to the large amount of SAS programming required. After evaluating the rest of the SAS/Intrnet package and other web enablement options, a decision was made to use a combination of Application Dispatcher programs, the SAS MDDB Report Viewer, and the SAS webEIS™ package. This combination provided maximum flexibility and the ability to rapidly generate new data marts as information was added to the data warehouse.

To aid in ease of use, the MDDB Report Viewer was modified to include some additional features. These included titles indicating the source of the data being viewed, as well as additional buttons on the report page that let the users view online help and the data dictionary as they were generating reports.

THE DATA WAREHOUSE INTERFACE

Since access to the data warehouse was to be delivered over the web, a web page was set up on the departmental server. This page contained links to public information about the University that was available to anyone, as well as basic information about the contents and accessibility of the data warehouse (see **Figure 1**). After the user entered a username and password, the SAS/Application Dispatcher was invoked to validate the user. If the username and password were valid, the user would be presented with a different data mart depending on their level of access. These data marts contain links to standard reports that are generated each semester by Institutional Research and
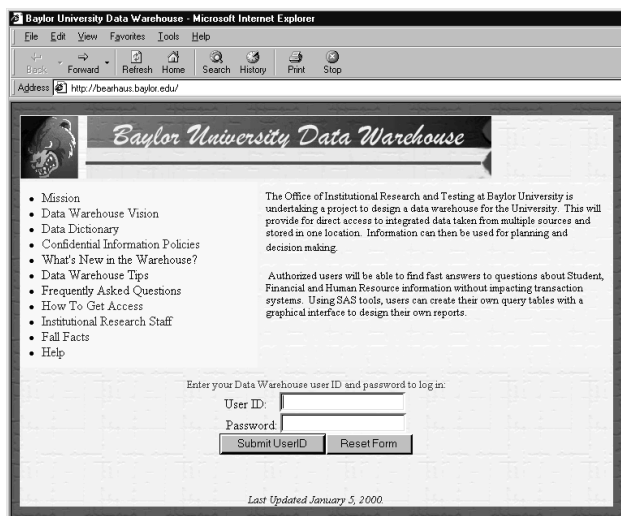

Figure 1: Main Data Warehouse Page

Testing, including enrollment counts and student profiles. Additionally, as users began to use the system and discovered new ways to view and use the data, other reports could be added to this page (see **Figure 2**).

The data marts also included access to Interactive Reports, via a link to the SAS MDDB Report Viewer. After clicking on the Interactive Reports link, the user is presented with a dynamically
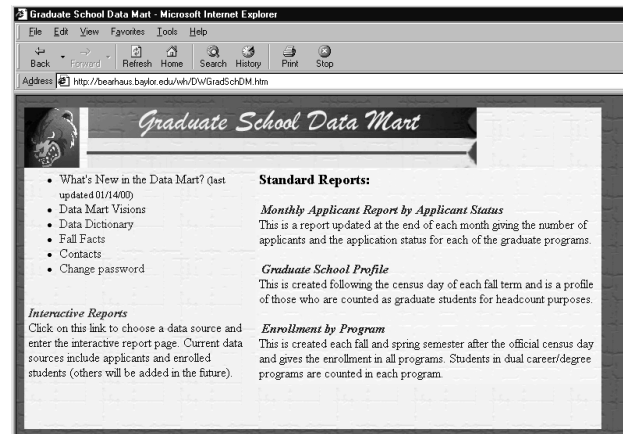

Figure 2: Graduate School Data Mart

generated list of data sources to which they have access. These data sources are the MDDBs generated by the Warehouse Administrator. The page is generated by a SAS program running through the SAS Application Dispatcher (see **Figure 3**). This page also contains some basic instructions on using the Report

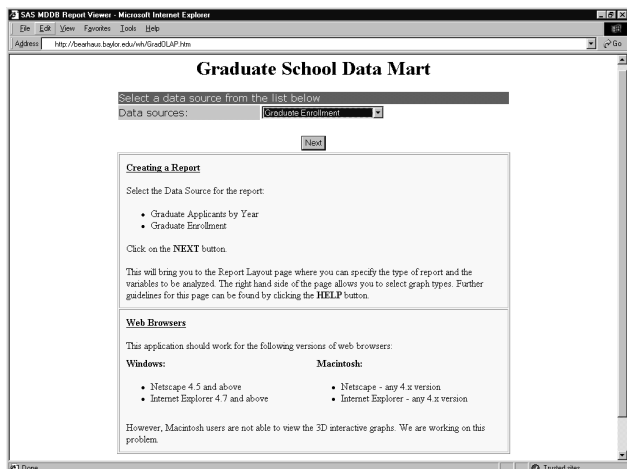Layout page which follows, as well as some notes about supported web browsers.

Figure 3: Data Source Selection

After a data source is chosen, the Report Layout page is dynamically generated and displayed. The Layout page contains information on the variables and statistics available in the data source. Included in the variable lists are the hierarchies defined in the MDDB. For student data, the main hierarchies are the Academic Year hierarchy, consisting of the Academic Year and Semester, and the Unit-Program hierarchy, which groups students by the Unit of the University (Business, Education, etc.) and their academic program (e.g., Bachelor of Science in Mathematics). It also allows the user to determine the appearance of the report page, including specifying a title and whether or not a graph of the data is displayed. This page contains links to a help screen and to the data dictionary, so that the user can look up what any particular variable represents (see
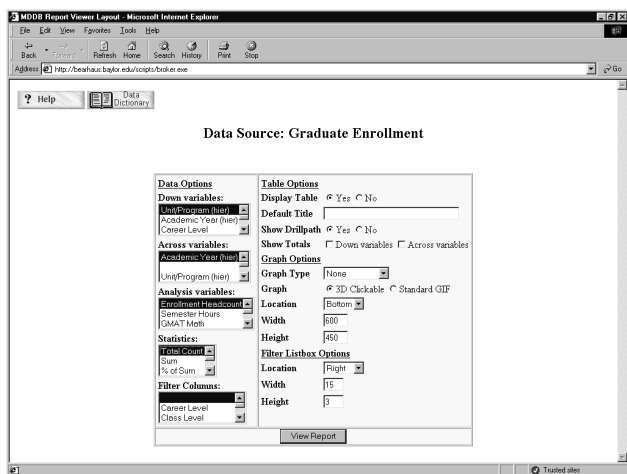

Figure 4: Report Layout Page

Figures 4 and **5**).

After the report has been laid out, the user sends a request to the Application Dispatcher to run a program to generate the report. This SCL program takes the information passed by the Application Broker (such as the metabase name and the row and column variables) and generates the report, which is returned as HTML to the user's web browser (see **Figure 6**). However, this report is not static. Users can change the row and column variables, or the statistic displayed, without returning to the layout
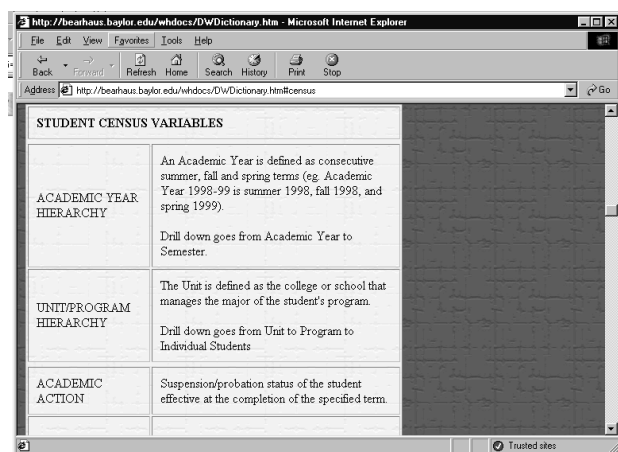

Figure 5: Data Dictionary

page. Other options include downloading the data into a spreadsheet, exchanging the rows and columns of the table, returning to the Report Layout page, viewing the online help, or viewing the Data Dictionary.

Additionally, if one of the variables displayed is a hierarchy, the user may drill down the hierarchy to a lower level. For example, the report in **Figure 6** contains hierarchies in both the row and columns. Clicking on the BUSINESS heading from the rows and
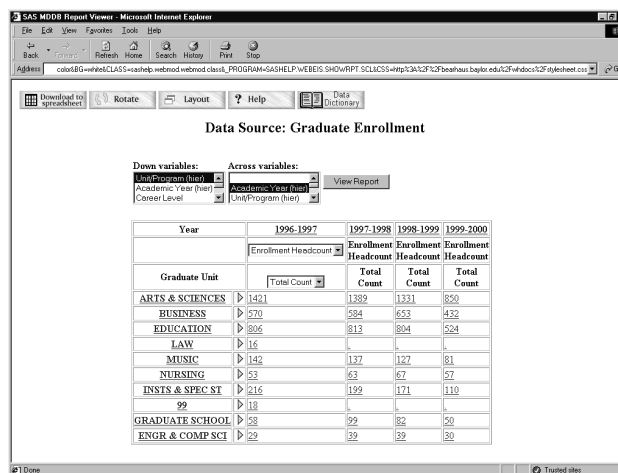

Figure 6: Report Page

then on the 1996-1997 heading from the columns leads to the report in **Figure 7**, which displays the enrollment counts for all programs in the School of Business for each semester of the 1996-1997 academic year. This information can also be presented graphically if the user has chosen to turn on the graphs. This feature makes use of the SAS-supplied GraphApplet that comes with the SAS/Intrnet package (see **Figure 8**). Some data marts also have access to Java applets developed in webEIS to view the data.

An additional feature is the ability to 'reach through' the MDDB and retrieve records from the underlying detail table. In any of the reports, clicking on a value in a table will bring up a window that allows the user to select the variables to display for each record and allows them to either display the records or download them to a spreadsheet. Another modification to the MDDB Report Viewer was made in this module to only retrieve records with non-missing analysis variable values. Access to the
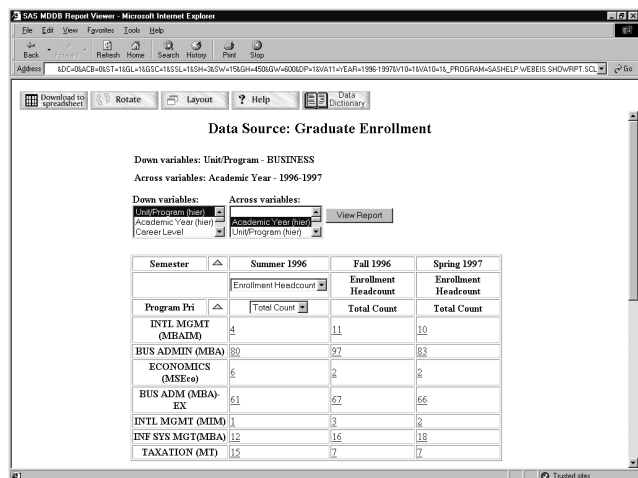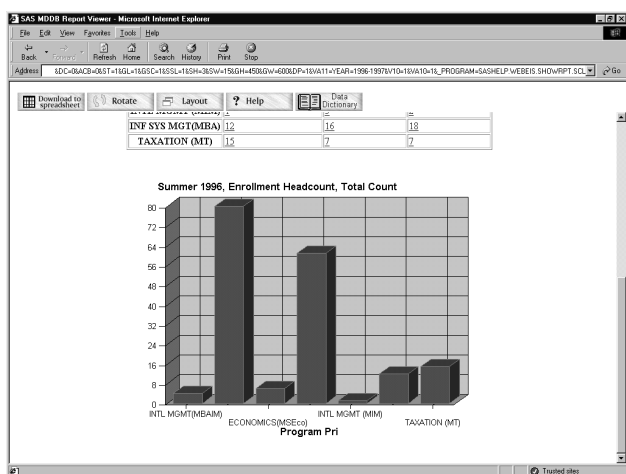
Figure 7: Report Page with Drilldown

Figure 8: Graphical Output

individual student records is restricted according to the user's access level.

OTHER APPLICATIONS

The Office of Admission Services is also using the data warehouse to aid in recruiting students. One tool allows the recruiters to access lists of students who have indicated an interest in the University. The list can be restricted by recruiter territory and areas within that territory, as well as by zip code, test scores, and the likelihood of the student enrolling in the University. A second tool (using webEIS) allows the director of recruiting to evaluate the effectiveness of the recruiters by comparing the number of students who have applied, been admitted, and paid a deposit to the goals set for each recruiter.

BENEFITS OF THE PROCESS

Several important benefits came out of the initial phases of the data warehousing process. First and foremost was greater access to information for decision making. This information was made available directly to administrators at their desks, and was quickly available to answer questions. This took some of the burden of ad hoc reporting off of the analysts of IRT and placed it in the hands of those who needed the information. Additionally, the turnaround time on the ad hoc reports was reduced from days or weeks to minutes or seconds.

A second benefit to the process was a greater emphasis on data cleanliness. As administrators began to access the data directly, they began to communicate the importance of clean data and error checking to the staff members responsible for data entry. It also led to a higher awareness of how 'dirty' some of the data was, even some of the archived data that had been cleansed already.

Finally, the process of developing the data warehouse and eliciting input from users led to a greater understanding of the needs of the those users. This should help the Office of Institutional Research and Testing continue to meet the information needs of University decision-makers, and to do so in a more effective manner.

CONCLUSION

In summary, providing access to a data warehouse through the web enabled the Office of Institutional Research and Testing to provide information more effectively to the administrative decision-makers of the University. By giving access to information directly to the end users, a significant savings in time was realized, both for the staff of IRT and for the administrators. This allowed decisions to be made more quickly and more effectively.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the authors at:

Phil Rhodes
Institutional Research and Testing
Baylor University
PO Box 97032
Waco, TX 76798
Work Phone: (254) 710-2061
Fax: (254) 710-2062
Email: Phillip_Rhodes@baylor.edu

Faron Kincheloe
Institutional Research and Testing
Baylor University
PO Box 97032
Waco, TX 76798
Work Phone: (254) 710-2061
Fax: (254) 710-2062
Email: Faron_Kincheloe@baylor.edu