

Data Warehouse Implementation: Where We Are 1 Year Later

Evangeline Collado, University of Central Florida, Orlando, FL

Linda S. Sullivan, University of Central Florida, Orlando, FL

ABSTRACT

There is a continuing demand across the higher education sector for increased extraction and analysis of university data to facilitate strategic decision-making and management of university programs and activities. This paper builds on last year's presentation and describes how data validation of imported warehouse data was completed, the conversion of existing SAS reporting applications to utilize warehouse data, and how additional SAS applications such as SAS Enterprise Guide®, Information Map Studio®, and Web Report Studio® have been integrated to provide the university user community access to strategic performance measurement, evaluation, forecasting and decision-making.

WHO WE ARE

MISSION

The mission of the Office of Institutional Research is to provide information that is timely, easily accessible and of the highest quality to enhance decision-making, strategic planning, and assessment at the University of Central Florida.

PURPOSE

Institutional Research (IR) provides electronic and web-based dissemination of official information to the University community (including the Board of Trustees, the various colleges, departments and other academic and administrative units), external agencies, and the Florida Board of Education (BOE); generates, supervises or develops all official University data reports and state-required reports; provides end-user data solutions and training; supports business intelligence analysis for decision-making purposes. The director and staff serve on numerous university-wide committees and workgroups and assist with the collection and interpretation of institutional data, assist in planning academic programs, and participate in the implementation of evaluative procedures. The functions of the office support the entire university enterprise.

BACKGROUND

Institutional Research has, as one of its primary responsibilities, the task of reporting all official data to internal and external constituents. Staff from the IR office routinely meet with end users and participate in workgroups and committees that deal with data needs and information access. This allows us the opportunity to interact directly with end users and determine, one-on-one, exactly what their data needs are and the format that would best meet their needs. This has the added benefit of allowing IR staff to more immediately respond to needs and make changes to our reporting environment as they are requested.

INTRODUCTION

At last year's SESUG conference, the UCF Office of Institutional Research presented a paper on the initial implementation of the university's data warehouse project. The focus of the presentation was on how the SAS® Enterprise ETL Server was used to import and merge more than 10 years and 5.5 million rows of student enrollment data from both legacy and ERP-system file structures as the initial data warehouse files.

This paper continues to chronicle the development of the data warehouse project during the last year. With the addition of 10 years' worth (just under 400,000 rows) of admissions files into the warehouse, we faced many of the same data quality and validation challenges as occurred with the student enrollment data. Three additional staff members were added to the project during the latter half of the year, providing the momentum which enabled the project to move forward in exciting ways.

The past year's development has focused on completing the data clean-up, which included converting a large number of records from legacy values to current student information system values and identifying and populating missing values; the establishment of a project stakeholders group representing a cross-section of the university users; the usage of several SAS® business intelligence (BI) client tools, by both programmers and power users, to develop pilot and production projects; conversion of existing programs that use SAS/IntrNet® software technology to utilize the client tools in the SAS® Enterprise BI Server suite and to retrieve data from the warehouse. The specific tools that will be addressed in this paper include SAS® Web Report Studio, SAS® Enterprise Guide®, and SAS® Information Map Studio.

DATA VALIDATION AND CLEANSING

The current files in the warehouse consist of official state-required census data. As loaded, these files contained only the historical data reporting elements mandated by the state. University reporting and analysis needs required

additional elements be added to the census files to obtain more robust output. Structured Query Language (SQL) queries were written to extract as much detail as possible from the student information system and load the new values to the files. Mapping tables and formats were used to create new fields when a one-to-one relationship existed between the existing code and descriptions. Data spanning 27 academic terms were pushed through this cleansing process.

After the initial updates were completed, data validation was necessary to identify where missing or incorrect values still existed. Frequency counts and summary tables were created to examine the data and locate invalid data groupings and mismatched or duplicate values. This enabled us to identify where weaknesses existed in the data and, after careful research, update the data files to be as complete as possible. Although very time-consuming, this process was extremely critical to ensure that the information extracted from the data warehouse is of the highest quality to support university decision making and strategic planning processes.

CONVERSION OF EXISTING SAS APPLICATIONS

Over the past five years, our office has developed several dynamic applications using the Application Dispatcher in SAS/IntrNet software to drive them. We will outline three of these applications and detail the conversion from the dispatcher to one of the new technologies offered with the SAS Enterprise BI Server tools.

ADMISSIONS PROFILE

The Admissions Profile provides our user community with application, acceptance, and enrollment statistics about new students to UCF via a web browser. The web display exhibits tables, charts, and maps offering the information needed to answer questions pertaining to counts, percentages, test scores, diversity, and geographic representation of the students on a term or annual basis. The dynamic display also provides trends and yields. The latest term's data is added when a new file is submitted to the state. We now have more than 10 years of admissions information.

The data update portion of this application required several steps and several large programs, written using SAS® software, to complete each term. Now that the data is in the warehouse, one scheduled extract, transform, and load (ETL) job is executed and the new term is appended. This has significantly reduced the time it takes to maintain this application. Figure 1 shows the admissions trends for fall terms over the past ten years as shown in the application using SAS/IntrNet technology.

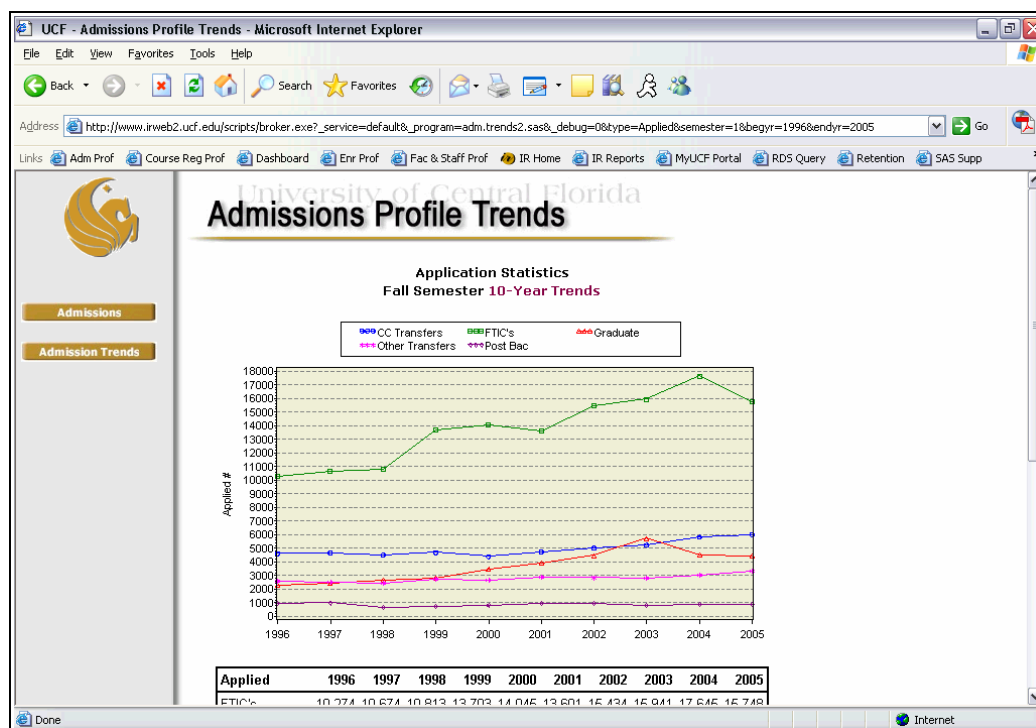


Figure 1

We decided to convert this application to a SAS® Stored Process that could be accessed through the SAS® Information Delivery Portal on the web. We wanted to improve the look and feel and provide the information in a different format. We chose to use SAS Enterprise Guide software to build the stored process. The drag-and-drop interface of this client tool makes application development easier, as the program code is written behind the scenes. Of course, we were able to modify the code to our specifications and save it within the stored process. But this

modification did not take the many hours that it took to develop the original application. The stored process begins with a user selection screen that is generated based on variables that are set as parameters. This eliminates the need to create a separate web form, based on either HTML or ASP language, for user input. Figure 2 shows trend information by region in a chart format and Figure 3 shows tabular results. This output shows acceptance and enrollment statistics, as well as the application statistics shown above, all on one web page. Modifications to the generated code were necessary to create the dynamic links for drill-down capability. The same code that was used in the original application to generate the dynamic links was used here with minor changes.

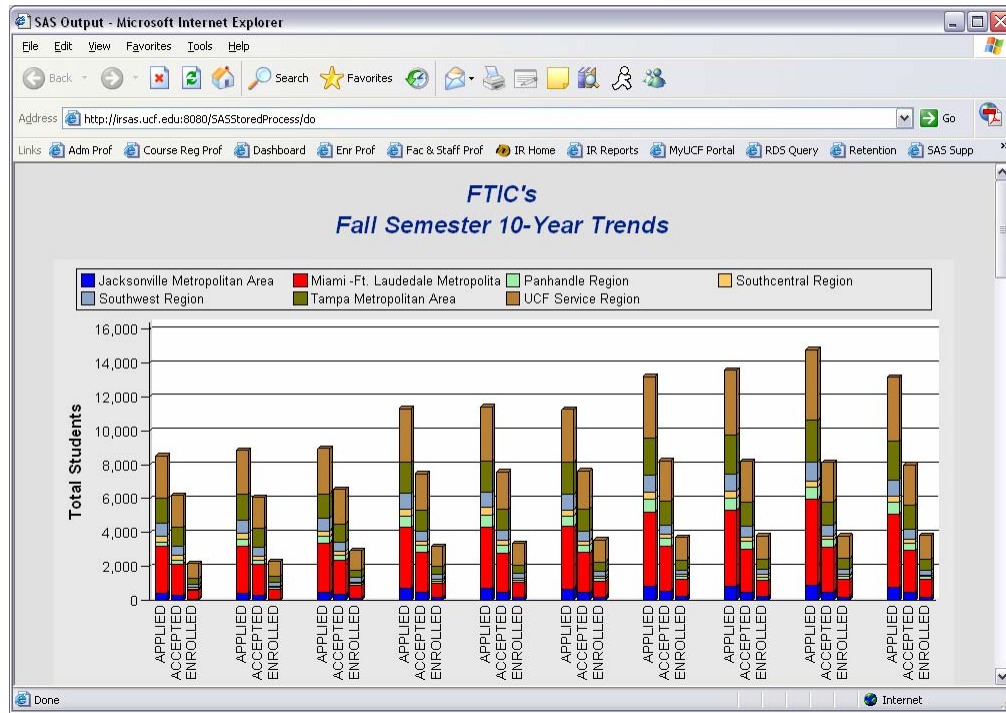


Figure 2

YEAR	REGION	APPLIED	ACCEPTED	ENROLLED	ACCEPTED-APPLIED RATIO	ENROLLED-ACCEPTED RATIO
1996	Jacksonville Metropolitan Area	426	287	84	67.37%	29.27%
	Miami-Ft. Lauderdale Metropolitan	2738	1871	516	68.33%	27.58%
	Panhandle Region	290	230	64	79.31%	27.83%
	Southcentral Region	356	264	99	74.16%	37.50%
	Southwest Region	722	545	173	75.48%	31.74%
	Tampa Metropolitan Area	1481	1095	354	73.94%	32.33%
	UCF Service Region	2500	1888	892	75.52%	47.25%
1997	Jacksonville Metropolitan Area	445	307	68	68.99%	22.15%
	Miami-Ft. Lauderdale Metropolitan	2763	1807	581	65.40%	32.15%
	Panhandle Region	391	273	92	69.82%	33.70%
	Southcentral Region	357	235	99	65.83%	42.13%
	Southwest Region	769	540	207	70.22%	38.33%
	Tampa Metropolitan Area	1566	1106	345	70.63%	31.19%
	UCF Service Region	2563	1820	892	71.01%	49.01%
1998	Jacksonville Metropolitan Area	461	331	99	71.80%	29.91%

Figure 3

ENROLLMENT PROFILE

The Enrollment Profile website was originally conceived as a way for the IR office to immediately respond to daily enrollment questions during key times of the year. The decision was made to utilize SAS/IntrNet technology to create a dynamic environment that provided information on student headcount enrollments in many different ways or “views.” For example, this application shows enrollments by college, undergraduate/graduate, full-time/part-time status, gender, ethnicity, classification and major in a drill-down fashion. This website was designed to replace and enhance numerous hard copy reports and be accessible via the web in a user-friendly, dynamic and interactive environment. Figures 4 and 5 show two of the views available in this application.

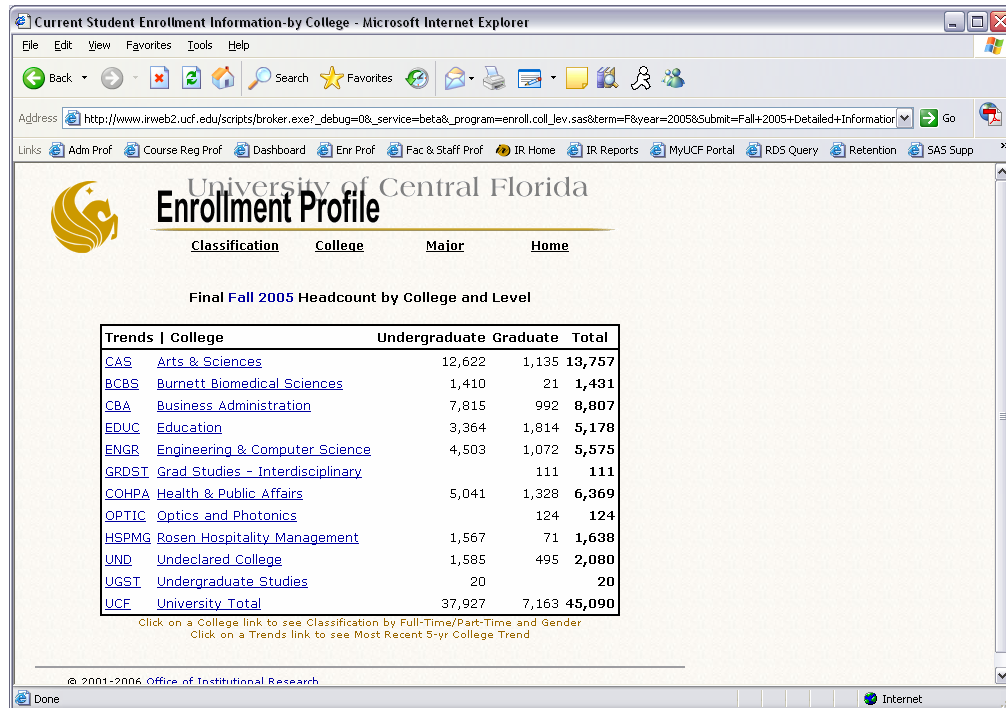


Figure 4

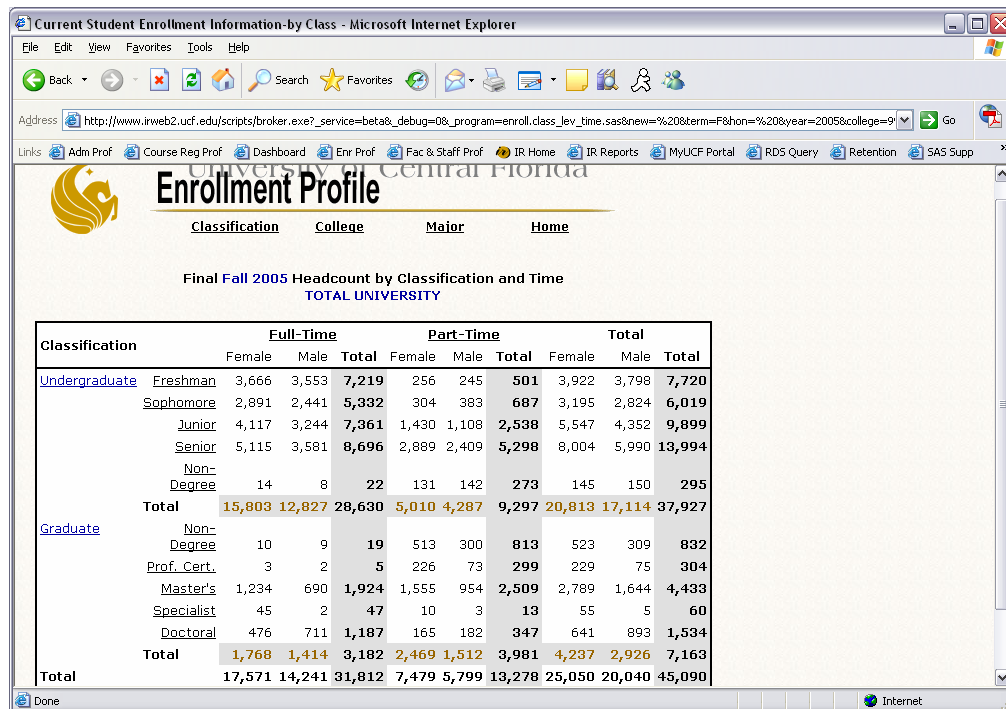


Figure 5

At full implementation, after considerable feedback by the user community, this application had more than 30 different tabular displays available by college or entire university. The users were provided with a mechanism to “slice-and-dice” enrollment headcount information in various ways. However, this functionality required the creation of numerous programs that had to be maintained over time.

We decided that the ideal environment for a tool with this capability was an OLAP cube. OLAP, which is an acronym for On Line Analytical Processing, is a mechanism to provide pre-summarized dimensional data, that meets UCF's business reporting requirements, in a user-friendly, drill-down environment with very fast response times. The information seekers are given the option to select how they wish to see the data, so they are free to create customized reports, instead of the IR office providing pre-designed views of the output and developing new programs every time a new view is needed.

SAS Enterprise BI Server suite provides a tool to create and maintain OLAP cubes in a user-friendly, simplified interface. Dimensions and levels of detail required to support our business processes were designed into logical hierarchies and the data was summarized at these levels and stored. Access to this information is provided by the following SAS BI client tools: SAS Enterprise Guide, SAS Web Report Studio, SAS Information Delivery Portal, and the SAS Web OLAP Viewer for Java and .NET. For more information on OLAP functionality provided by SAS, please visit <http://www.sas.com/technologies/bi/olap/index.html>.

RETENTION AND PROGRESSION PROFILE

The university community needed a way to track how many students returned for classes each successive fall term; i.e., how many students were retained by UCF. Green bar reports had been generated by the Information Technology – Computer Services Department, but these were very cumbersome to read. Users who needed the data in an electronic format for further analysis would have to manually enter this information into the client tool on their computer.

Institutional Research developed a database where networked users can map to retrieve the data needed for their analysis. However, this solution did not satisfy the “everyday” user who just needed the numbers. An application was developed using SAS/IntrNet technology to provide two different reports, one that contained detailed information for the most recent cohort years, and another that provided statistics for a particular cohort year. Figure 6 shows an example of the web display that gives a user the opportunity to select the type of report. The user is next prompted for various selections including what format he/she would like the generated report – PDF or MS Excel. Figure 7 provides an example of a generated PDF report.

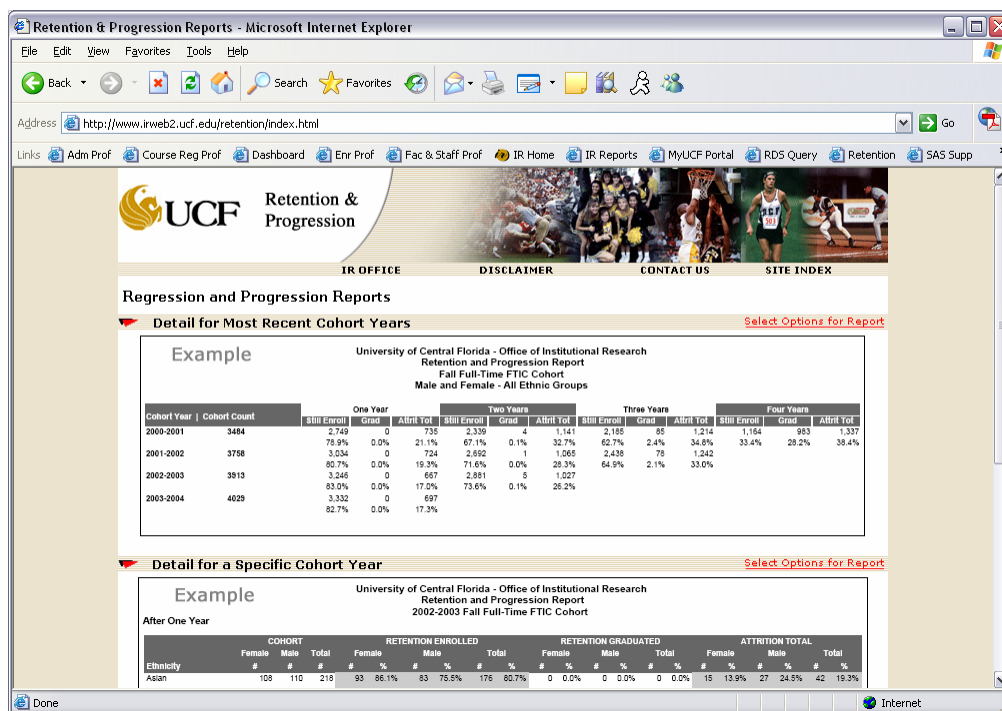


Figure 6

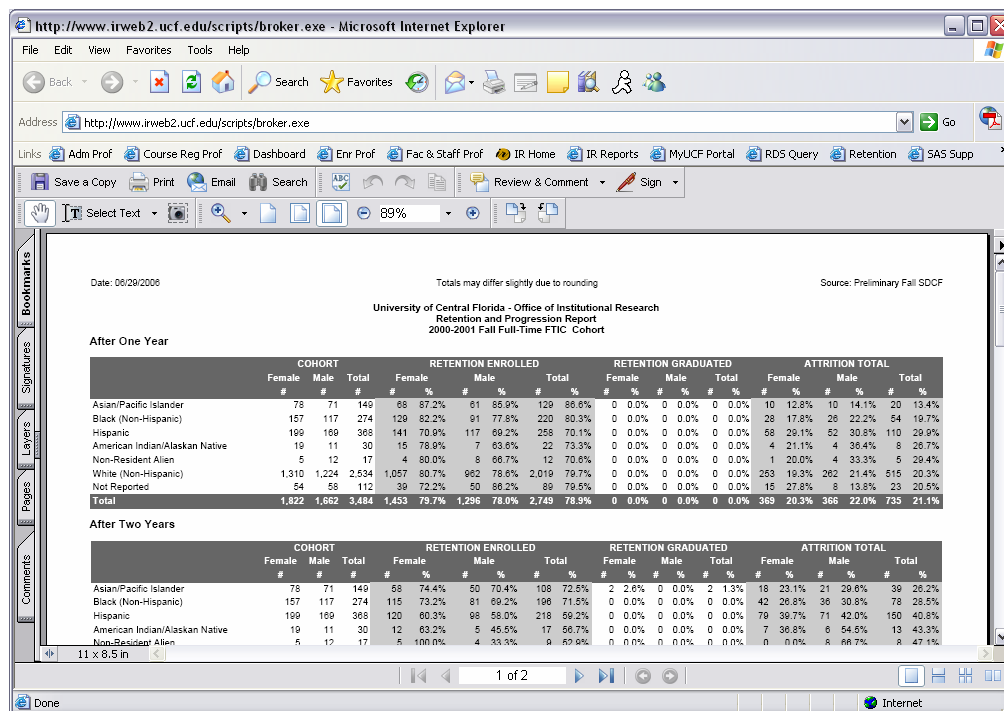


Figure 7

The logical progression of this type of application was to convert it to a stored process that would generate the reports after the user is prompted for the parameters. The stored process can then be accessed and executed in MS Excel, using the SAS® Add-In for Microsoft Office client tool, for those users who need to have the information in this format. For others in the user community who prefer a web report, this stored process can be added to a portlet in the customizable SAS Information Delivery Portal and executed on demand. The data for the stored process would come from the data warehouse, where a retention fact table was created, which would eliminate the need to have users map to a network drive. All cohort years are now in one table in the warehouse, whereas, each year had to be in a separate table in the database software we used prior to this implementation. This makes querying the data much easier.

INTEGRATION OF SAS® BI CLIENT TOOLS BY POWER USERS

A highlight of project development this year was the ease with which an IR Office power user with no SAS software programming skills was able to convert a major university reporting database from Microsoft Access over to the data warehouse using Enterprise Guide, Information Map Studio, and Web Report Studio. When the project in Enterprise Guide was complete, an information map was created which pointed to the required data elements, and a report was developed to display the output of the project using Web Report Studio software. This project demonstrates how SAS BI client tools enable the user community to independently create reports. The staff member found the tools user-friendly and was able to customize the generated code to meet specific needs of the project.

Figures 8, 9, and 10 are examples of the project.

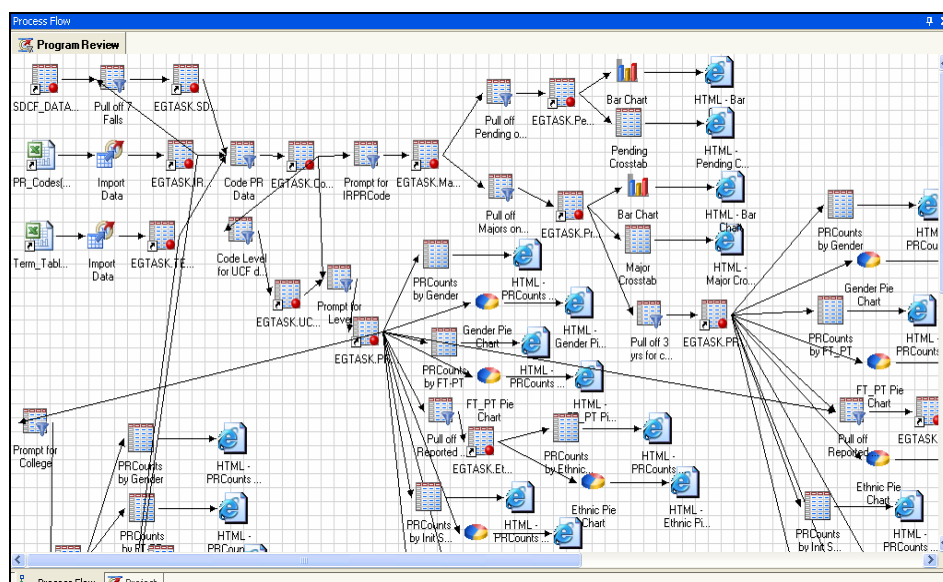


Figure 8

		GRAD			UGRD	Total degrees at all levels
		Doctoral	Masters	Specialist	Bachelor	
1969-70		.	.	.	430	430
1970-71		.	19	.	736	755
1971-72		.	150	.	1187	1337
1972-73		.	184	.	1504	1688
1973-74		.	198	.	1540	1738
1974-75		.	308	.	1532	1840
1975-76		.	323	.	1745	2068
1976-77		.	378	.	1752	2130
1977-78		.	343	.	1981	2324
1978-79		.	354	.	1989	2343
1979-80		.	303	.	2085	2388
1980-81		1	309	.	2134	2444
1981-82		.	285	.	1830	2115

Figure 9

Degrees Granted by Academic Term											
						2004-05					200
						1210	1220	1230	Acad. Year	1240	1250
						Summer 2004	Fall 2004	Spring 2005		Summer 2005	Fall 2005
GRAD	Doctoral	BCBS	Biomolecular Science PhD		PHD	1	1	.	2	.	.
				Total		1	1	.	2	.	.
				Total for College at degree level		1	1	.	2	.	.
	CBA	Business Admin PhD	Accounting Track	PHD	.	.	2	2	1	.	
			Finance Track	PHD	1	.	.	1	1	1	
			Management Track	PHD	.	.	.	.	.	.	
			Marketing Track	PHD	.	.	1	1	.	1	
			Mgmt Info Systems Track	PHD	.	.	.	.	.	3	

Figure 10

CONCLUSION / FUTURE PROJECT DEVELOPMENT PLAN

During the coming year, concurrent with completing the conversion of the remaining programs based on SAS/IntrNet software, the project will begin to expand beyond a census structure. Selected data files from Finance and Accounting and Human Resources will be added to the warehouse, as well as institutional survey data and the Faculty Activity Reporting System. Additionally, a data mart interface will be developed between our Reporting Database Service (RDS) database, which contains current and future academic year information, and the data warehouse. This will enable the user community to generate robust trend analyses from historical through current and future term data.

The seamless working interface of the SAS® Data Integration Server (formerly SAS® ETL) and the SAS Enterprise BI Server has enabled this project to move forward quickly with limited staffing and skill sets. The reporting capabilities and output have been enthusiastically received by the university user community. This next year will see continued progress and development in providing the university user community with access to strategic performance measurement, evaluation, forecasting and decision-making through SAS BI client tools and the data warehouse.

CONTACT INFORMATION

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

Evangeline Collado
Office of Institutional Research
University of Central Florida
P.O. Box 160021
Orlando, FL 32816-0021
Work Phone: (407) 823-5061
Fax: (407) 823-4769
Email: ecollado@mail.ucf.edu
Web: <http://www.iroffice.ucf.edu>

Linda Sullivan
Office of Institutional Research
University of Central Florida
P.O. Box 160021
Orlando, FL 32816-0021
Work Phone: (407) 823-6683
Fax: (407) 823-4769
Email: lindas@mail.ucf.edu
Web: <http://www.iroffice.ucf.edu>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.