

Banner Conversion, SAS Converted
Utilizing SAS to Reform our Data and Reporting Structure
Alan Socha and Keith Stiles

Required implementation of SCT's Banner student information system at UNC system schools has lead to increasing demands on Institutional Research offices involved in critical data collection for required reporting functions. But beyond crucial reporting functions, there is a growing need for critical business intelligence surrounding data for campus decision-makers. The Office of Institutional Research at Western recognized that implementation of Banner along with the implementation of CollegeNet's Series25 products represented an opportunity to better serve the decision-making processes by incorporating non-system related reports and data tools aimed more at providing reliable and critical data for campus users. This paper focuses on Western Carolina University's approach and experiences with data collection, cleanup, and reporting for the 2006 fall term including information concerning approaches to incorporating data discrepancy checks (and the accompanying cultural changes), potential pitfalls of implementation and reporting, improved business processes, and communication issues.

Why Choose SAS?

Prior to Banner conversion, data extraction for required reporting functions was managed through a series of COBOL-based programs that created flat files for posting to the General Administration servers (primarily FRED over the past couple of years). The majority of these programs were a decade or more old and had received irregular patching to deal with changing coding structures and needs. Application programmers in the Information Technology division were responsible for coding changes to these programs in consultation with Institutional Research. These split responsibilities

produced any number of difficulties surrounding data reporting due to the need to coordinate all program and coding changes through two offices. Over time, relics of earlier systems were lost in increasingly arcane programs not fully understood by current staff in either office.

In May, 2006, OIRP began moving forward with development of new data extraction programs using SAS. Preparation of our data framework caused us to assess the relative pros and cons of using the Alliance PERL scripts or charting our own course. Perl, a programming language originally developed for text manipulation illustrates its utility mostly in web development, network programming, and GUI development. It obviously has a wide range of quite valid uses. However, Western had no expert PERL developers and adopting these programs would have required someone on campus, either in applications development or in OIRP, to learn the programming language. We chose not to invest increasingly scarce resources to develop PERL programming skills.

We chose to develop our systems in SAS, a business intelligence program. SAS was chosen primarily because of its higher utility over PERL in data management and higher-order statistical analysis. Also guiding our decision to use SAS was the fact that we were already using it for reporting and data analysis. It has native ability to import and to export data in almost any format and can be linked to Microsoft Office, as well as many different types of databases. Using SAS seemed the most logical choice for importing data from both Oracle and Resource25 for the creation of our data files. This choice also allowed us to leverage our knowledge of SAS to develop the extract programs and simultaneously use those programs as a springboard to a more robust discrepancy process allowing us to move beyond the required reporting duties for the system. The

Alliance provided copies of their PERL extract programs that we used as base for locating data elements in Oracle. From that point, development of these programs took the remaining time between May and the September census date to complete.

There were a number of important reasons for our choice of SAS over PERL and other alternatives. First, developing these programs in SAS integrated well within the existing reporting structure required by General Administration. As we all know, SAS is used exclusively on the FRED server for producing both the error frequency reports and NCHED/IPEDS reports. It was a logical extension for us to develop our own processes in SAS for maximal leveraging of our discrepancy process on campus. The fact that we happened to have a talented SAS developer of course aided this choice!

Second, SAS integrates well with all available software tools commonly in use on campus including Banner (based on an Oracle database framework), Resource25 (facilities inventory, academic and event scheduling software running on Microsoft SQL Server 2003), and Microsoft Office (Access and Excel). The fact that SAS was designed over 26 years ago as a tool for data analysis, data manipulation, and data reporting fit perfectly into our campus with the move from the mainframe environment to a PC-centric environment. The ability to extract data analysis, report on it, and present it in formats familiar to faculty and staff made the decision far easier.

Third, SAS is an overall better data management tool than PERL or any other alternatives available to our campus community. SAS supports SQL and other development tools more readily than any other possibilities available. Most of our pre-existing SAS reports imported flat files. Adopting SAS as our development tool allowed us to both build our flat files and to expand on this reporting function to also build a more

stable and efficient datamart. Our program structure allows us to create data files, perform internal discrepancy checks on the data and build this datamart as SAS datasets. These datasets are based on the flat data files we submit each semester, but may be modified to include additional Banner variables as needed. Banner IDs and campus codes are just a couple of examples of these extra variables. The inclusion of Banner IDs serves to facilitate matching our datamart back to Oracle for analysis of additional variables not in the data files. Their inclusion also facilitates updates like Social Security number changes. Similarly, the campus code allows us to track our students taking courses at other institutions (i.e., UNCA, the community college, Penland, etc.).

Fourth, SAS allowed us to improve our overall internal data collection and enhance correct coding of data on the front-end while creating the requisite flat files for reporting. As you will see, we used the required reporting elements as a base from which to delve deeper into our campus data looking beyond normal coding problems and looking for data that raises flags or concerns. A great example of this is using non-reporting discrepancies to identify pieces of data that simply do not match (i.e., an employee title that does not match the accompanying function code or department code). Additionally, SAS's easy integration as a data management tool means that users may check their discrepancy reports long before the census date relieving data cleanup issues post-census and highlighting areas of concern much earlier than was ever possible before.

Finally, adopting SAS as our data management tool allows OIRP to be self-reliant. While operating in SIS+, OIRP was reliant on application development programs from the applications programming staff in the Information Technology division, one of many users that they needed to serve. Moving to SAS allowed the office to be less reliant

on the division for applications development and freed them to work more closely with the office in correcting base Banner data issues, issues that could only be modified with more fine grain control and security than that granted to functional areas or to OIRP.

The Data Framework

Western's data framework consists of a group of SAS programs that create our flat files and build discrepancy reports for the functional areas to examine. In some cases, these reports are available through our reports portal page and can be run at a moments notice. Other reports due to their nature are e-mailed to the functional areas for examination. Figure 1 outlines a broad overview of the program structure (the shaded boxes are still under development).

Running this sequence of programs is simplified through the efficient use of macros. These macros are used to determine which files are created, where they are saved, and the term for which they are being created. Macros are also used to set the username and password for connecting to both the Oracle and Resource25 databases allowing anyone with access to those database to run these programs. These macros are all set in the *Flat File Creation & Discrepancy* program, the only program the user needs to open and run. This program calls a *Formats* program which imports lists of valid codes for discrepancy checks. For example, CEEB codes, FICE codes, CIP codes, and county codes are all imported from an Access database using this program. The Access database allows for easy maintenance of these codes. The code updates were initially built manually from the PDF documentation on the FRED website. General Administration has now posted SAS catalogs of these codes to the FRED server. By using an external validation, we may check our actual Banner validation tables to ensure

that they are valid and up-to-date. We will be modifying our *Formats* program to pull from these datasets removing manual updating from the process.

The program user chooses which files run simply by commenting or removing comments for a few lines of code. The student courseload file and course grade file are both submitted with a course description table file. Each of those programs calls the program that creates and runs discrepancy checks for the course description table file. The *Flat File Creation & Discrepancy* program also contains macro variables that set whether or not the programs will be run for update files. These are used to update our tenth day files and correct errors for final submission to General Administration.

Also included in the program structure is a Social Security number change program. This program matches the student data files back to Banner by Banner ID to determine if any Social Security number changes have taken place. Then, this program creates the Social Security number change file for submission as well as updating all of the SAS datasets in the datamart with the Social Security number changes.

Each data file that is created is checked for discrepancies. The discrepancy reports that are created are very similar to those on General Administration's FRED website. This is more convenient than continuously posting files to FRED to be checked, and also allows for additional internal discrepancy checks. Some of the data files and their discrepancies are also exported to an Access database. This database has queries that use the data files and Banner data to display more specific error reports allowing quicker and easier error correction process. In particular, we can display original Banner data prior to the coding for the official frozen files, basically decoded data.

OIRP's decision to create the data files using SAS was generally a positive experience. It allowed the office to become more involved in the Banner conversion process. Examining the data enabled the office to tie the functional areas together and improve the integrity of the institution's overall data. This process also enhanced OIRP's role as the center of data reporting on campus. The only disadvantages we have really encountered are that some of the SQL code has to be modified in different programs which can be a bit inefficient. Also, there is little to no support from Information Technology on SAS given the relative few number of users on campus. However, our overall experience has illustrated to us that we made the right choice for our campus in moving to a SAS-run data framework.

The data framework allowed us to leverage data collected for reporting for integral decision-making processes and other related reporting duties. A number of reports were created as spin-offs of this data framework including a budget report, additional enrollment and retention forecasting models, an Asheville enrollment report, Fact Book reports, and retention and graduation rate reports. Processes for these reports are now almost completely automated. With the press of a button, our office can run a suite of reports. On the census date, we create frozen data files, send the Administration and Finance Division a budget report, and send our office of WCU Programs in Asheville an Asheville enrollment report. The commonly shared data files also enable us to produce data for our Fact Book at the same time. Due to the increased efficiencies, the simplicity of importing datasets, and the ease of linking datasets together, data forecasts, budget projections, and ad hoc reporting have become easier to manage while becoming far more precise and accurate.

Incorporating Discrepancy Checking

Traditionally, data discrepancy at Western merely meant correcting data entry and coding errors identified by error reports for General Administration reporting. Functional areas never really understood the relevance of the census date and discrepancy checking to themselves and to the larger institution. Discrepancy checking only gained relevance after the census date and immediately prior to reporting deadlines. We were not really managing our data but reacting to it. In developing this process, OIRP wanted to move the campus beyond looking for mere data entry and coding issues. We wanted the campus to think more strategically about all data collected and analyzed on a daily basis. We also wanted to frontload data correction giving data consumers the opportunity to look beyond the surface level of their data.

Data discrepancies now go beyond just the likely issues that are encountered on the GA reporting files. Allowing users to examine their data earlier and in more detail aids in producing cleaner, more reliable data. Cleaner data directly correlates with better and more reliable data analysis and forecasting, integral functions in the increasingly competitive market for new student enrollment. The discrepancy reports allow checking of any number of items of interest including a teacher certification list, contract courses, student teaching courses, distance learning courses without faculty or enrollment, residential living residency counts, courses with invalid total faculty load, resident courses without a room assignment, a faculty member, or enrollment, courses with invalid start and end times, and many more.

Any number of potential pitfalls can occur while implementing new data systems and a new data discrepancy process. Chief among these potential problems is not

completely understanding the new data systems. Where does Banner store particular data elements? What tables store integral data required in reporting data to the system office? How did the campus structure shared data elements? These are just some of a myriad of questions faced in developing a discrepancy process in a new system.

A second and closely tied pitfall is communication. It is integral to maintain clear, constant, and consistent communication among all users, potential and actual, of campus data. Information needs to flow freely between functional areas, information technology, and Institutional Research. Each of these large groups has an integral role to play in the development of a robust system that meets the basic requirements of data reporting in a state system. Demographic data is obviously the easiest part of the process. But, even demographic data can be problematic where all potential data users have not been informed or consulted on data conversion and import procedures. Yet, even more integral to basic reporting functions is the structure of shared validation tables and shared data elements. These can cripple data analysis quickly. Beyond these issues, there lies the issue of looking at student data or personnel data from a more strategic position, looking beyond what we need to do today and thinking about what we may need a year or to later as it pertains to coding. Open lines of communication are key to raising the awareness of users in functional areas that data is more than just fields in a form or entries in a table; data has consequences and bad data can harm the institution.

A third pitfall surrounding discrepancy processes concerns the importance of the census date. As mentioned earlier, at Western, there really was not an urgency surrounding data cleanup until near or, more normally, after the census date. The functional areas were missing the potential loss of student headcount or student credit

hours along with the accompanying budget dollars and faculty positions. Adding a new discrepancy process with a new student information system has increased awareness of the potential issues surrounding the census date and its importance to all campus constituencies, not just the Institutional Research “bean-counters”. Users in the functional areas are now realizing that being late, or lagging behind the census date, can and will harm their interests whether that be in budget dollars or faculty positions. And, as a few learned recently, pushing data entry to the last minute can cause higher blood pressure!

Finally, the largest pitfall that you may encounter while attempting to implement a new discrepancy process is the campus culture. Books have been written about organizational culture, both what it is and how to change it. One of the largest initial impediments encountered at Western was, and, in many respects, still is, changing the “this is how we’ve always done it” attitude. Many years of operating in SIS+ and working with procedures implemented for that system created comfort with processes that were difficult for many areas to part with as they moved into Banner, including in Institutional Research. Our goal is to place Institutional Research at the center of data management linking the functional areas together. The discrepancy process serves as a good starting point for discussions of meta-data issues and ways in which all functional areas can work together to strategically engage data issues and how data can be used to aid decision-making.

Talking About Business Processes, Data Standards, and Data Stewardship

One of the more important elements of developing a successful data discrepancy process is the need to clearly describe campus business processes, not only to verbally

describe them but to get them in writing. Documenting your business processes has some very immediate positive impacts on data reliability and usefulness. One perennial problem area for Western has been teacher education, both degree-seeking and licensure only students. As one of our first steps towards improving our data structures, OIRP convened meetings including College of Education representatives, the graduate school, and undergraduate admissions. The initial meeting centered around figuring out proper coding for the ubiquitous GA “Y” and “C” flags for teacher licensure. Banner has no native location for storing this data which meant the campus needed to identify an alternative method to collect it. Searching for a solution to a coding problem lead to a realization that we were in need of detailed analysis of our coding structures and business processes surrounding degree-seeking education students and their licensure only counterparts (Capturing Data on Education/Teacher Certification Students: Business Processes for Collecting and Reporting Education and Teacher Certification Students in Banner, 2007).

Over the next few months, the functional areas, the College of Education, and OIRP worked to refine a clearly outlined business process document aimed at standardizing practices concerning data entry on teacher education students that would more accurately capture integral data surrounding the number of students in each area and their status in the professional education sequence. Refinements to these processes are ongoing with the business process document being, in many respects, a living document. The traditional student process was outlined, described in detail, and flowcharted. The group decided that the required “Y” (for students in a teacher licensure program) and “C” (for students who have completed degree requirements or licensure

requirements in a teacher licensure program) flags reported to General Administration would be captured through use of the non-CAPP requirement fields in Banner. Placing this data in this particular location gave the College the ability to track removals from the program, reinstatements, all current students in the pipeline, and completers. This data was not readily available to the College prior to Banner implementation. SIS+ solutions for providing this data were erratic and unreliable at best.

Engaging in this business process definition for teacher education students yielded numerous positive improvements that were leveraged at both the undergraduate and graduate levels. Discussions concerning initial coding of student admit types and student type coding revealed the need for a standardized system to deal with these coding issues outside of teacher education students only. As a result, the committee produced and shared a coding matrix that integrated these two related Banner fields with undergraduate admissions and the graduate school. Adoption of the coding matrix was quickly completed giving both functional areas a guide to aid proper coding of students at admission saving later efforts needed to correct and clean up miscoded student records.

This process also brought resolution on where to store and develop General Administration's required 12-digit CIP codes. Prior to Banner implementation, the CIP codes for specialty teaching areas were hard coded into the flat files each semester representing an inefficient use of staff time and requiring ongoing manual coding that never ended up being stored in SIS+. In consultation with the Registrar, the Banner form, SOAXREF, was identified as the storage location for the 12-digit CIP codes. OIRP maintains the official list of CIP codes based on General Administration's Academic Program Inventory of approved programs for the campus. Program codes from

SMAPRLE are mapped to the 12-digit CIP codes with a thirteenth digit used to identify distance education only programs.

The bottom line of engaging in only one documentation of a business process was the ultimate clean-up of multiple issues affecting data entry and coding on campus. As a result, our teacher education reporting and analysis functions have never been more robust. At the same time, difficulties being experienced with graduate teaching students were identified and largely resolved. And, the admit type coding matrix yielded improvements in student coding beyond education students. Business process discussions are now a more integral part of any campus discussion surrounding student data coding and entry; and, these discussions are yielding more strategic examinations of why we code data in one fashion or another. Development and discussions are now also examining financial aid, finance, and personnel coding. In many respects, this process is one of three that helps us ensure that functional areas consider the impacts of coding changes that will affect all users of campus data.

In order to have business processes that work, you have to have quality data. As a part of this overall new procedure and approach to campus data, OIRP instituted a Data Standards committee charged with improving data quality on campus. Data quality (including deployed reports) affects almost all areas of a campus's business processes and operational/strategic decision-making. Data is the foundation of all information systems whether they be student, financial, or human resources related. Clearly, the functional areas and specified individuals on a campus have responsibility for entering and maintaining the data. However, OIRP, at least at Western, functions as a partner with these areas and seeks to facilitate and support commitments, processes, and actions at all

levels of administration to ensure responsible data stewardship of the university's data and systems.

The data standards committee meets regularly to examine data entry standards and has produced significant improvements in communication among functional areas since its inception. The committee oversees all changes to Banner codes, especially those codes dealing with shared validation tables. Modifications to shared validation table codes in Banner must be communicated to the committee. The committee then examines the changes looking for potential impacts of changes to multiple areas. Once the changes are thoroughly examined, the committee approves the modifications and the codes are added. For non-shared validation tables, the committee requests that the functional areas communicate their proposed changes for informational purposes. Those changes are then shared with the larger campus community who may use that data for reporting or analysis purposes.

The committee has also made significant progress in removing unused entries from validation tables. Many validation tables were provided by SCT pre-populated and contain a host of codes that the campus simply does not use or need. The committee has been working through these tables, and, in some instances, has already managed to half the size of some of them. The rule of thumb is that if a code is available someone will use it! The committee has also produced a procedure for flagging duplicate records. They are now in the process of developing methods for the removal of duplicate student entries from the system. Work is also continuing to modify and update the data standards manual. The work of the committee has immeasurably improved the quality of campus data and decision-makers ability to make decisions based on campus data.

Heavily associated with the need for data standards is the need for data stewardship. Where do the responsibilities for each of the pieces of our data system reside? Who owns data? What are the liabilities associated with data stewardship? The purpose of our Data Security and Stewardship Committee is:

1. To protect the university's data and to protect the university from misuse of its data.
2. To provide a framework defining the appropriate protection required for each broad category of data;
3. To define how categories of data are established and assigned; and,
4. To define who is responsible for ensuring that data is handled in an appropriate manner (Data Security and Stewardship Policy, 2007).

The policy applies to all university enterprise-level data, whether or not it is centrally managed. Enterprise-level data is defined as data that the University has responsibility for, or is critical for the operation of the University (Data Security and Stewardship Policy, 2007). The policy is also defined in such a fashion as to cover all forms of data, electronic records, paper records, and data housed on or off campus (Data Security and Stewardship Policy, 2007).

The Data Security and Stewardship Policy at Western clearly delineates the role of Institutional Research in maintaining data integrity. OIRP "has primary responsibility for meeting the University's reporting obligations and overseeing the movement of unit record data between the campus and the University of North Carolina. It is the responsibility of all other divisions/units charged with the reporting of institutional data to ensure that OIRP has a record of the parameters of such reporting and timelines that

OIRP will maintain as part of an inventory updated annually” (2007). The Director of Institutional Research serves as a member of this committee.

This new policy classifies data into four different groups: confidential, third party confidential, internal, and public. Obviously, access to confidential data is limited to staff that “need the data to perform their job functions” (Data Security and Stewardship Policy, 2007). Each level of access below confidential opens the data to more users/viewers and is based on specific procedures for handling data set by both legal statute and by campus policies. Periodic reviews of classification requirements are conducted with concomitant changes occurring as necessary. As part of this policy, all staff authorized to “access or disclose Confidential or Internal information are required to sign a confidentiality statement each year as part of their evaluation process” (Data Security and Stewardship Policy, 2007).

The three elements of business process documentation, data standards, and data stewardship complement each other to increase the integrity and the trustworthiness of the data entered into campus information systems. They each represent one piece of an overall discrepancy process that not only identifies potential problems but then codifies and maintains the structure of the data systems to ensure accurate, timely, and reliable information to decision-makers. Engaging the university community in these types of discussions and activities brings the various stakeholders together and assigns accountability for potential problems should they arise. As we all know, it is far easier to pass the proverbial buck if no one knows who is accountable for the data.

Conclusions

[Insert conclusions here]

References

Office of Institutional Research and Planning. Western Carolina University.
*Capturing Data on Education/Teacher Certification Students: Business Processes for
Collecting and Reporting Education and Teacher Certification Students in Banner.*
February 20, 2007.

Western Carolina University. *Data Security and Stewardship Policy.* February
26, 2007.

Office of Institutional Research and Planning. *Data & Reports web page.*
<http://planning.wcu.edu/data.htm>.

Figure 1
Data Framework

