NASCIO 2016 State IT Recognition Awards

Title: Student Success Analytics Project

Category: Data, Information, and Knowledge Management

State: Georgia

Contact: Angela Bell
angela.bell@usg.edu
404.962.3069

Project Initiation Date: June 1, 2014

Project Completion Date: December 1, 2015

EXECUTIVE SUMMARY

The importance of higher education – whether at a college, university, or a technical school – to individual success and the state's economic competitiveness is at the center of Complete College Georgia, an initiative sponsored by Governor Nathan Deal.

The University System of Georgia sought to support the goals of Complete College Georgia by strengthening its efforts to recruit undergraduate students from a broad range of high schools and to provide the support services necessary to ensure successful completion of their higher education.

To strengthen their efforts, university system leaders turned to data. They already collected vast amounts of data. But what did the data actually tell them about student recruitment and enrollment patterns, the communities from which students come, how students move through college, and their chances of success in higher education?

University system leaders recognized the need for new and innovative tools and techniques for gaining greater insights into their data and using those insights to inform decision making on policy and operational matters. They also recognized the need to merge their data with outside sources.

The resulting **Student Success Analytics Project (SSAP)** brought together experts from various disciplines and created groundbreaking analytics for data-driven decisions **in all 30 of the university system's colleges and universities**. The project addressed a series of defined challenges, and it transformed data to make longitudinal analysis possible. The use of data visualization represented another innovative approach for extending analytic capabilities beyond more conventional reporting techniques.

In one example of the innovative approaches made possible by the project, university system leaders are able to view the student pipeline from high schools to higher education as social networks and to identify state colleges and universities with similar recruitment and enrollment patterns. Being able to identify high schools that are only loosely affiliated with state colleges and universities leads to the development of strategies for engaging them and thereby helps expand student access to higher education. In another example, data visualization illustrates patterns in how students enter and move through college. Campus leaders are able to identify students in need of support services that help prevent them from dropping out of school.

The project brought **sophisticated analytic** and **business-intelligence capabilities** to the USG's central office and its colleges and universities and is helping to equalize the ability to **make data-driven decisions across the entire university system**. It demonstrates the power of embedding data and analytics in decision making to understand and predict student behavior, drive more effective allocation of scarce financial and human resources, and ensure better educational outcomes.

CONCEPT

University systems and institutions of higher education nationwide are under growing pressure to measure their performance. Increasing the number of graduates and accelerating time-to-degree are both critical components of **Complete College Georgia**, an initiative sponsored by Governor Nathan Deal.

By 2020, over 60 percent of jobs in Georgia will require a certificate, associate's degree, or bachelor's degree. At present about 42 percent of the state's young adults are prepared to such a level. To remain economically competitive, Georgia must not only maintain current graduation levels but also produce an additional 250,000 graduates in upcoming years.

To help guide decisions affecting the state's college-completion efforts, the leadership of the University System of Georgia (USG) realized they needed new and innovative tools and techniques to gain insights from the multiple gigabytes of data collected by colleges, universities, and the USG's central office. These tools and techniques could facilitate data analytics, business intelligence, data visualization, and data-driven decision making.

The USG faced significant challenges in integrating analytic capabilities and implementing a data-delivery infrastructure. The **Student Success Analytics Project (SSAP)** successfully tackled those challenges and created groundbreaking analytics for data-driven decision making at the USG central office and the university system's 30 colleges and universities. The SSAP brought together subject matter experts from a range of disciplinary backgrounds, innovative analysis, and integrated data tools. It culminated in actionable intelligence on student enrollment patterns, pathways through college, and predictors of success.

To achieve innovative analysis and data visualization, project leaders formed a partnership involving the USG's Division of Research and Policy Analysis, the **Division of Business Intelligence**, and the Carl Vinson Institute of Government at the University of Georgia. While the USG collects gigabytes of data for federal reporting and institutional research, the data were gathered on a term-by-term basis and not transformed for longitudinal analysis.

At the onset, the SSAP partnership identified and set out to solve the following problems:

Data storage and accessibility: Data were stored in numerous tables by academic term, which did not facilitate longitudinal tracking, a necessity in student-outcomes research.

Lack of connections to external data sources: Understanding student-enrollment patterns requires knowledge of the secondary schools and communities from which students come. It also requires linking the USG's data to other relevant data sets.

Need for diverse and cutting-edge analytical approaches to education data: While traditional statistical approaches answer many questions, increasing stratification and mobility in higher education and the positioning of the USG's institutions in a changing environment require the perspectives of demography, geography, social network analysis, and big data.

Need for data visualization for analysis and presentation: Conventional data reporting in tabular formats limits the information that can be relayed to decision makers simultaneously, powerfully, and succinctly. Data visualization itself was identified as a way to achieve more interactive data inquiry.

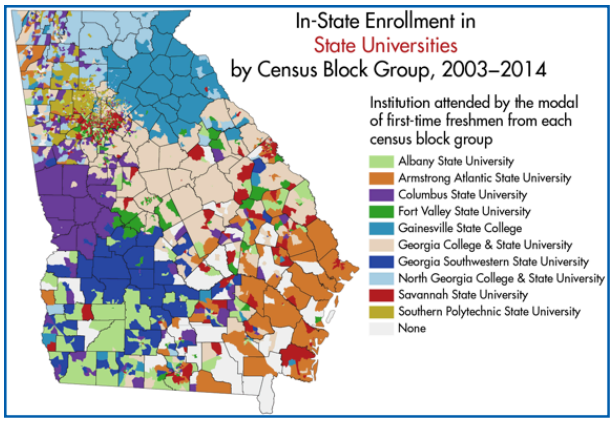
SIGNIFICANCE

IT innovations began with the development of a longitudinal database that integrated data from disparate term-by-term USG data tables. Restructuring the data into one row per student per academic term created a longitudinal database that could be used for analysis over time. The SSAP then appended external data, primarily from the National Student Clearinghouse and the U.S. Census Bureau, by geocoding student addresses. **The resulting database included nine million enrollment records covering 12 years and an estimated 2,800 Census Bureau variables.** With input from USG subject matter experts, cutting-edge analytic techniques were applied to the transformed data. These techniques, which originate in multiple disciplinary backgrounds and use advanced computing and statistics, revolutionized the university system's research and analysis.

For example, taking advantage of the different disciplinary backgrounds of the subject matter experts working on the SSAP, USG leadership was able to view the student pipeline from high school to a USG college or university as a social network. They focused on the number and intensity of connections between the state's various high schools and USG colleges and universities. They studied how the network of feeder high schools changed over time, which colleges and universities have similar networks and why, and how the pipeline from high schools to colleges and universities is a dynamic and malleable relationship. Their analysis enables them to identify high schools that are only loosely affiliated with USG colleges and universities.

As a result, effective strategies can be developed to engage those high schools and thereby expand student access to higher education.

Incorporating **Census Bureau** variables provides even **greater analytical depth** to the development of these engagement strategies. The Census Bureau variables are about the communities from which students come and not about income and housing of individual students. USG research shows that the community from which students come matters.
The percentage of the population with a bachelor's degree in a student's community is a strong predictor of college completion. Additional, more comprehensive analysis showed that the percentage of men in the community with manufacturing jobs is also related to bachelor's degree completion, but the association moves in the opposite direction; the prevalence of jobs not requiring completion of a bachelor's degree serves as a disincentive to college completion. By appending Census Bureau variables to the USG data, leadership has information that enables them to tailor their messages for particular communities and ensure more effective recruitment.



In an example of data visualization, USG data about first-time freshmen is integrated with data from the U.S. Census Bureau to depict enrollment patterns by census blocks.

Additional innovation involved implementing modern visualization software to dynamically illustrate patterns in how students enter and move through college. For example, data scientists adapted an interactive cord diagram to show international migration patterns and depict students who attend multiple institutions. Visualization tools add a means of analysis that's difficult to accomplish with traditional reporting tools, and they meet Gartner recommendations to use a diversity of tools to accomplish standard and exploratory reporting.

The SSAP clearly highlights two important process improvements for the USG:

- The project developed and documented the method for turning disparate data tables into a comprehensive, integrated longitudinal data set complemented by important outside data sources. The longitudinal data set can now be easily maintained and updated with current data collections. It ensures consistency in definitions across analytic projects and eliminates the weeks of time that have historically been required to gather, merge, and organize data for individual research projects.

- The project developed a model for assessing analytic needs across the university system and addressing those needs in a collaborative and efficient way.

The SSAP identified common needs and requirements and connected talented data scientists and business-intelligence tools. It completes the loop from campus data collection to collective innovation and back to the decision makers on campus. By pooling data in standard formats and connecting them to outside data sources, campuses avoid unnecessary duplication, and equivalent capabilities are assured across both large and small institutions of higher education.

IMPACT

The SSAP brought analytic capacity to the USG's central office and actionable intelligence to decision makers. Some of the analytic tools were in such immediate demand by campuses that they were deployed through a web portal.

As a result, the USG is enhancing educational policy and operational practices to improve service to students.

- **Student enrollment patterns –** Analysis of enrollment patterns refers not only to a student's initial enrollment in the system but to patterns for the next 18 academic terms after initial enrollment. Pattern analysis and visualization create a string of enrollment states, such as EE-EE-AEEEEEGGGGGG. It shows a student enrolled two terms (EE), then took a summer off (-), repeated that pattern, transferred to another system institution at the beginning of the third year (A), and then enrolled continuously for another five terms before graduating in the fall of the fifth year (G). Aggregating these patterns over the entire system enables the development of probabilities at certain points in time for whether a student is likely to drop out. The vast amount of information from the USG's data about the backgrounds of students and other enrollment behaviors – number of hours, major, and term and cumulative grade point average – allows further refinement of those probabilities, and the information can be used to target additional support to students at risk of dropping out. Targeted student-support interventions range from supplemental financial aid to additional advisement and career guidance. **Further data analysis will allow USG leadership to determine the most effective dollar amounts for financial aid and scholarships** and thereby **promote enrollment and retention** while at the same time reaching the greatest number of students possible. Data models are also being used to estimate the effect of policy and practice changes on enrollment.

- **Pathways through college –** An early finding from the enrollment pattern analysis was that the pathway through higher education for many students who earn a bachelor's degree in four or five years involves attending another system institution at some point as a transient student. Think of the student who takes a course in organic chemistry at a college near his or her home during the summer. While it allows the student to give full attention to a difficult course, it opens up the possibility for a student to take a course that doesn't fulfill a requirement at the home institution. **Bolstered by findings from the SSAP, a system-wide policy committee** is now assessing ways to **eliminate barriers to enrollment** as a transient student while ensuring students have the information necessary to enroll in the right courses.

- **Predictors of success –** A pilot program is currently under way at Georgia Gwinnett College to predict the success of newly enrolled students and target support services to students with lower probabilities of success. The program uses the university system's historical data, including student demographic characteristics and academic preparation; Census Bureau data about students' home communities; and their college record to develop probabilities of success. While similar efforts are taking place at many colleges and universities across the country, the USG's approach is unique. The project emanates at the **system level and enhances the accuracy of the predictive models through the use of system-wide data and external data sources**. Furthermore, the advanced data science technique of a random forest model is better able to employ a vast number of predictive variables than typical statistical techniques.

The transformation of **data** into an **analysis-ready format**, the development of analytic tools, and the addition of an interactive business-intelligence interface puts information in the hands of USG policymakers more quickly and enables them to base their decisions and planning on more sophisticated and predictive analysis. While some campuses already had advanced business-intelligence capabilities, many others struggled to move beyond required reporting. As a result, the **SSAP has enhanced analytic capabilities at schools where they were lacking and is helping to equalize the ability to make data-driven decisions across the entire university system**.

Another example of data visualization depicts intra-system transfers between state universities (blue), comprehensive universities (gold), research universities (pink), and state colleges (green).

Financial savings and cost avoidance are another important component of the SSAP. Technologies shared across the university system increase savings, and informed decisions can also reduce costs. In addition, the model of shared analytics is repeatable and can be embedded in annual decisions, which drive future savings.

The SSAP saves money by having created an analysis-ready data set with appended data from external sources. The USG estimates that it would take one FTE four months per school to recreate the data set for one institution, which totals 120 months of data preparation across the entire university system. The Carl Vinson Institute of Government at the University of Georgia developed the longitudinal database in nine months with one FTE. In addition, enterprise contracts for professional services, software, and hardware reduce costs even further. **Consistency across all USG campuses simplifies and constrains costs for future technology updates**.

The SSAP's predictive and prescriptive capabilities lead to cost-effective decisions. Uninformed decisions can result in higher costs through incorrect predictions of demand for housing or academic majors or by continuing an expensive educational service that is not effective or no longer needed.

The SSAP demonstrates the current and future power of embedding data and analytics in decision making to understand and predict student behavior, improve educational services, ensure greater student success, and effectively allocate limited financial and human resources.