Doing Academic Analytics Right: Intelligent Answers to Simple Questions

Ganesan (Ravi) Ravishanker, Wellesley College
Overview

It is difficult to imagine an argument being made against data-driven decision making. It might sound something like, “Let’s ignore the evidence and make our strategic decisions based on anecdotes and guesses.” As ridiculous as that might sound, the fact is that many important decisions in higher education are made this way, simply because harvesting the right data to inform our decisions is more complex than it might at first appear.

Some apparently simple questions yield either stunningly complex answers or a surprising number of requests for clarification. For example, ask a higher education registrar, “How many students do we have?” and you are likely to hear, “It depends. Do you want the number of full-time students only, or the combined number of full-time and part-time students, or the number of full-time equivalent students? Should the number include graduate as well as undergraduate students, non-matriculated as well as matriculated students, students in continuing studies and/or summer institutes?” And on, and on…

Higher education senior administrators ask many fundamental questions for which the answers are not straightforward. And yet, in the wake of the recent financial crisis and continually changing landscape in higher education, data-driven decision making is ever-more essential. Frankly, it is not enough to simply know how many students we have—we need to know and be able to refresh answers to questions about retention rates, resource consumption by major, and so forth. Once we understand these, we need to notch up a level and engage in modeling and forecasting.

What motivates an institution to do what it takes to answer the new, important questions that drive strategic decisions? What dynamics are involved? Who are the players? How do we get beyond positioning academic analytics (the higher education equivalent of business intelligence) as just another project to complete or just another engine to build? If we agree that creative thinking, consensus building, and agility are some of the elements that distinguish strategic from tactical thinking, what can information technology (IT) do to enable a culture that is capable of making sound, strategic decisions based on dynamic interactions with truly informative data?

Many institutions have invested large sums in the promised nirvana of enterprise resource planning (ERP) systems, and they are frustrated that even years after ERP implementation, they do not have access to reliable data. Why?

ERP systems are excellent transactional systems that help us capture critical institutional data and store it in relational databases. Unfortunately, for the most part, ERP systems do not provide the much-needed infrastructure for reporting or for keeping some forms of data history. Because of the severe resource crunch in higher education, the implementation of the ERP is often the focus, and, despite strong recommendations from IT professionals, provisioning for reporting becomes an afterthought. It is only after the ERP goes live that people begin to appreciate that it’s much easier to put data in than get it out. Soon the institution realizes the urgency of the problem; after all, what good is having the data if we can’t use it to inform important short-term and long-range decisions? In response, individual departments with tech-savvy staff begin building their own mini data marts and keeping their own history, effectively growing a silo reporting enterprise.
In parallel, IT staff members begin to create a reporting database and enterprise reporting infrastructure for reports that are typically run several times a day and are stored for retrieval by the appropriate faculty and staff. Report names, as well as the naming of content elements, are often inconsistent, and eventually these become a source of constant confusion. Soon, the community realizes that this is not an efficient method. Even the smallest change to a report requires IT staff intervention, and after an ERP upgrade, many of these reports have to be rewritten. In addition, depending on how these reports are written, you may not have access to semester data until the semester rolls over!

An absence of institutional data definitions complicates matters further, and reports catering to different functional offices might have different definitions of the same variable. When senior administrators sit down to look at reports, each vice president brings data that doesn’t reconcile, and valuable time is lost in minutiae about why the data results are different. Quite simply, the promised land of ERP quickly turns to mud.

Thankfully, as a result of recent advances in data warehousing and dimensional data modeling, we can find solutions to these problems. As we have witnessed numerous times over past years, however, technology alone cannot be the solution to all problems. In this ECAR research bulletin I explore the various factors that must come together for an institution to have an academic analytics infrastructure that is flexible, agile, appropriately structured, and cost-effective. I examine not only appropriate technologies but, more importantly, the critical roles that stakeholders and governance play in setting the stage for success. As we know, there is a tremendous amount at stake for American higher education right now. How can IT facilitate—and motivate—our institutions to adopt the tools that can help mine the gold that resides in our very own vaults?

**Highlights**

No doubt the scenario explained above is far more common than one may think. In a detailed ECAR study conducted in 2005, an overwhelming percentage (over 65%) of respondents said that they were in the process of implementing a better reporting infrastructure. Given that the implementation of a data warehouse solution for better reporting is resource-intensive, and the fact that much of higher education was unprepared for the financial crisis in 2008, I doubt very much that many institutions have made much progress. In addition, “Higher ed usually follows about 10 years after business adoption.”

**Making the Case for Analytics**

When I came to my previous institution, Pace University, and spoke with the faculty and staff, it was abundantly clear that having access to data was their primary concern. “Get us the most accurate and timely data as quickly as possible,” they said. As a result of previous administrative decisions that had been made, and also because of the data silos that had developed, sharing of information between schools or functional offices was minimal, and faculty members wanted this changed. In other words, we needed to democratize the data access while, of course, operating within the confines of data integrity, privacy, and security.
Pace had purchased a software solution to the reporting problem, especially for the senior administration. The community was excited by the possibilities the software provided because some of the IT staff were doing excellent work to make the software work. Unfortunately, no analytics solution can work without a few key ingredients such as common data definitions, a solid data dictionary, a strong data model, and a data warehouse (for both speed and history). Thanks to agreement on some of the data definitions among the leaders, the software dashboards were becoming very useful. However, reconciliation continued to be a problem, and most importantly, the IT staff were struggling to keep up with the changes and limitations of the software. For example, semester rollovers were tedious to deal with.

As a result of this early attempt at a homegrown analytics effort, the senior administrators of the institution became keenly aware of the weakness of the whole reporting infrastructure. Everyone began to understand why we have the issues we have. We took this opportunity to present to the Administrative Systems Steering Committee and then the president a plan to resolve the reporting issue. Our plan was to look for a modern, extensible, and functionally rich data warehouse solution. Since Pace, like many institutions, does not have a single ERP system that covers all functions, we needed a solution that allowed us to integrate data from diverse sources. We also outlined the process needed to implement it. By clearly explaining the capabilities of ERP and what it can and cannot do, and by providing the reasons why we needed a budget and roadmap for a strong data warehouse implementation (several months for each module), we were able to convince the senior administration to sign off on the project proposal.

For resource-intensive projects such as these, support from the senior administration is clearly essential. At Pace, we were able to garner support from the president who, along with the senior administrators, needed a lot more data to make data-driven decisions. By providing specific examples of how the new systems would expose the data and empower end users to access it, we were able to gain support from a varied group of users.

**Enterprise Reporting: Invest, Monitor, and Document**

Now let us look at the key ingredients for having a successful reporting infrastructure—the backbone of academic analytics.

After careful research for a data warehouse solution based on the criteria explained above, and after following the good practice of talking to many other colleges and universities, Pace chose iStrategy solutions. Our choice was based on what we considered to be sound knowledge of the higher education space, an excellent understanding of the dimensional data modeling, and the Kimball methodology for data warehouse implementation. Finally, this product also comes with literally hundreds of reports that work with Microsoft Business Intelligence for Higher Education and integrate well with Microsoft Active Directory for security management. In addition, the infrastructure allowed for easy creation of dashboards with drilldown capabilities. Since Microsoft products are commonplace in our environment, users are likely to feel comfortable with the delivered reports. We purchased the student, finance, human resources (HR), and financial aid modules, to be implemented in that order.
As with all large projects, process is extremely important, and identifying the stakeholders is key. Although IT organizations often have several analysts who are also subject-matter experts, their expertise is very different from that of individuals who work in the functional offices. In talking to institutions such as Rensselaer Polytechnic Institute and the University of Richmond, we learned that we needed to partner with Institutional Research on this project for it to be successful. We arranged for the vendor to provide several demonstrations for various constituent groups, including more detailed sessions for the director of Institutional Research and the president.

Upon the approval of all involved, we created the implementation team consisting of members from all key functional areas and IT staff. In addition, it was made clear that when we worked on specific modules, there would be increased participation from the relevant functional areas. We also prepared a governance and conflict resolution structure, which is essential.

Led by the Office of Planning, Assessment, and Institutional Research (OPAIR), Pace has been working for the past three years on a data integrity and data definitions document. This has become an excellent resource and starting point for us. In other words, we have a head start in this area.

One of the things that iStrategy did was load the Pace data into the data warehouse, which took two days to complete. During the initial meeting of the implementation committee, the actual data, including all of the pivot tables and several reports that were close to the hearts of functional staff, were shown with actual Pace data. This made the meeting lively, and attendees began to understand the monumental task in front of them while being pleased that now they had access to data at a level that they had never had before.

The definitional conflicts arose in the usual areas. What defines a full-time student? How should one deal with double majors in terms of departmental and faculty load analyses? How should one deal with missing data? Should financial aid calculations be made based on the initial statement by the student regarding his/her campus preference or the current choice of the campus? Should continuing degree program students (BA/MA) be treated as undergraduates or graduate students? How do we do cost of attendance calculations when a student takes classes at more than one campus?

This is where the expertise of Institutional Research comes in handy, and partnering with them is critical to the process. IT staff will, through years of experience, have significant contributions to make in resolving these conflicts, but they cannot and should not assume the final responsibility for these definitions.

As is evident from the Pace example, a very strong data warehouse and an academic analytics software suite are essential in this world of data-driven decision making. We have also seen that the other key ingredients necessary for a successful implementation are strong senior administration support, common institutional data integrity standards and data definitions, a clear governance structure, and a well-defined budget and implementation plan. Of course these ingredients are the same for all major projects we all engage in, but it is better to revisit them, learn from past mistakes, and refine them for success. For instance, having strong senior administration support typically results in
setting a realistic implementation timeline, which in turn can be used to drive the implementation plan.

**What It Means to Higher Education**

It is obvious that higher education is changing. Regardless of which institution you are in, the pressure to adapt to changes around us is weighing heavily on all of us. It is critically important in this environment to use all the information at our disposal to make informed decisions. Yes, data alone cannot and should not be the driving force in certain cases, but in a majority of the cases, it can tell us a lot. It is also true that skillful combining of data elements is what produces valuable information; data elements themselves are simply building blocks. For example, researchers understand that while collecting data about “how many” or “how often” is crucial to the research endeavor, the raw results are just the starting point: the true value of the research is the result of knowing which data elements to combine and compare in order to produce new knowledge. Interpreting and analyzing data is a highly specialized skill—we must not assume that simply exposing raw data results in good decision making. As counterintuitive as it might sound, data warehouse specialists advise that there is no such thing as “the data.” There is, however, data that is useful to an individual or group, so the point is to find it and make the best use of it.

In the corporate sector, data analytics has advanced tremendously, and many businesses live or die by it. There are many examples of how real-time data drives many of the important business decisions. In the book *Super Crunchers: Why Thinking-By-Numbers Is the New Way to Be Smart*, Ian Ayres discusses many examples of this. Specifically, he talks about how using sales data from stores in the areas that were affected by hurricanes, Wal-Mart was able to rush strawberry Pop-Tarts (because they do not require refrigeration or cooking and “people would be yearning for the gooey comfort of this”) to Florida stores that were in the path of Hurricane Ivan in 2004—prior to the arrival of the hurricane.

In another example, Ayres talks about how Harrah’s casino used terabytes of gambling data collected from gamblers’ swipe cards. Harrah’s calculates a “pain point” for each gambler, based on historical information. This is the loss that a gambler is willing to accept and still consider the experience to be enjoyable enough to want to come back. Harrah’s computer systems evaluate how the gambler’s current stint at the casino is proceeding, and when the gambler approaches his or her pain point, a “luck ambassador” is dispatched to ease the pain away by providing the gambler with a free meal at the steakhouse (again, a data-driven choice based on past visits).

Increasingly, businesses are using real-time social networking trend analyses to improve their business or for crisis control. Almost all major corporations now have social media response teams that mine data on social media using software that uses advanced natural-language processing to identify and respond to the “buzz.” If a negative sentiment is building up, immediate response by the people at the company is called for. For example, Comcast noticed many disgruntled customers complaining on Twitter about Comcast’s service, and Comcast now has a response plan almost in real time that avoids the snowballing of such bad press. In “How Companies Can Use Sentiment
Analysis to Improve Their Business,” Maria Ogneva, director of social media at Biz360.com, outlines various ways in which companies can take advantage of the emerging field of sentiment analysis with customer relationship management (CRM). It is well known that companies such as Amazon use historical data from purchase records to make real-time recommendations for related books or accessories that might interest you. They also use e-mail to suggest new arrivals that you might enjoy.

**Getting Started with Academic Analytics**

Data analytics for higher education is growing. Whereas a handful of institutions use data analytics to drive business decisions, it is fair to say that higher education is still at a stage of cautious adoption. For example, Purdue University’s Signals is an excellent data-mining application that mines student data from multiple sources to produce “signals” that are then used by various offices to intervene early and help students succeed. In the 2008 *EDUCAUSE Review* article “Action Analytics: Measuring and Improving Performance That Matters in Higher Education,” Donald Norris et al. discuss some of the major reasons that higher education might be lagging in academic analytics: academic culture favoring analysis over action, reputation over the need to improve academic performance, and autonomy over accountability.

Norris and his coauthors also provide examples of effective academic analytics being used in both for-profit institutions such as University of Phoenix and in nonprofit institutions such as the University of Alabama and Coppin State University. Given the vast amount of data at our disposal, it is easy to argue that the current use of data analytics in higher education is a drop in the bucket of what is possible. As Norris and others point out, the primary obstacles to greater use of analytics are resources and culture, and they are intertwined. Competition for resources is very much tied to the culture (should we spend hundreds of thousands of dollars for business intelligence software, or should we hire more faculty to teach?), and it is complex (the administrative definition of data is very different from my needs, so how can I be sure that I can get what I need to do my work?).

It is abundantly clear that we collect huge amounts of data that we are not using. It can be used to improve efficiencies, increase enrollment, boost retention, and improve teaching and learning. In order to get started, we certainly need the resource commitment from the senior administration, but we also need some creative thinking, consensus building, and agility.

**Enrollment, Retention, and Financial Aid**

For those who have taken the plunge, depending on the stage of the academic analytics process an institution is in, the questions for which we seek answers will vary. For example, when one begins the process, some of the most interesting questions are likely to be related to enrollment, retention, and financial aid. As the process matures, questions will shift to finance (including fundraising) and then to human resources. On the other hand, others may start with a relatively straightforward implementation, such as finance, to show to the community the value of academic analytics in bringing efficiencies, easy access to data, and decision making.
Enrollment trend analysis and setting appropriate goals for the upcoming years are among the most important and necessary activities that many of our institutions engage in. Having access to reliable and consistent data that are scientifically interpreted by trained professionals will give an institution a tremendous advantage. A team consisting of staff from Institutional Research who are trained in statistical methodologies, functional office staff with data analytics skills or training in informatics, and IT staff who are engaged in enterprise reporting will provide this advantage. A team like this, for example, along with staff from enrollment management, can perform multivariate analyses of easily accessible and well-defined historical institutional data that will result in a deep understanding of the profile of our students and help target others to improve our enrollment goals. Many institutions already have processes in place to do this, but unfortunately, only a few do this in a way that is strategic and inclusive.

Similarly, various functional offices on any campus are engaged in providing student services and therefore are influential in student retention. Many of them do various types of assessments, the results of which are fractured and not centralized. By providing an infrastructure to bring all of the data collected to a single repository, the knowledge the institution can gain is tremendous. This, for example, is repeatedly demonstrated at the Retention Advisory Board meetings at Pace, where we learn in every meeting about a new data source with partial information, which, when combined with some of the other collected data, provides an important new perspective.

Questions relating to finance are among the most important and relevant questions in the current environment. What is the true cost of attendance? Before we get ahead of ourselves, we certainly need a canonical definition of cost of attendance, and clarifying this definition is critically important for setting tuition and discount rates. With the appropriate analytics in place, we can assess how different discount rates (and resulting enrollment) would impact the institutional budget.

**Learning Management Systems, Classroom Usage, Network Flow, and Web Analytics**

Many institutions collect significant amounts of other information that they don’t analyze. Learning management systems (LMSs) hold a wealth of information, although it is often difficult to access without appropriate policies and IT intervention. What are the student and faculty access patterns, how many artifacts are associated with a course, how are students and faculty using the system, and so forth?

Similarly, we have untapped data on classroom usage, network flow, classroom technology use and failure rates, web server access, and more. Imagine if the provost had easy access to the LMS usage metrics: how many classes are using the LMS and in what ways (is it just for posting syllabi or for active use with blogs, wikis, and discussion forums), or what types of classes are using LMS (is it primarily the sciences, or humanities; is it used for larger classes or smaller ones as well?). This information, combined with other information such as grades, course evaluations, and retention rates, can be very helpful to understand if the LMS is helpful in learning and, if so, whether it can be used more widely in the institution. Similarly, classroom and other facilities usage...
can be analyzed to see if there are other revenue-generating opportunities, such as renting classrooms or technology access to other organizations in the area.

Web analytics\(^{10}\) can help explain whether we are getting our money’s worth from our investments in web services. Appropriately capturing the total cost of producing a website (including staff time) and measuring that against the web traffic and goals (such as increase in volume of gifts and donations to the institution, or the number of e-mails from a web form) can provide glimpses of the effectiveness and geographical reach of our web. Are we attracting enough traffic from international markets, and if not, does it pay to develop a portal for international visitors and prospective students?

**Managing the ROI**

The irony of successfully implementing academic analytics to improve decision making is that without careful management, the project can grow very expensive very quickly. Fortunately, by applying the right principles from the start, costs can be managed by viewing academic analytics as an investment, monitoring report usage, documenting definitions, and demonstrating value. Vigilant application of the principles of analytics to the process of enterprise reporting is a key to success. In this way, a “self-sustaining” dynamic project continues only as long as there is demonstrated value. As long as the institution interacts dynamically with enterprise reports and the intelligence they provide, the return on investment can grow well beyond the initial cost of the project. In the same way that institutional data can inform good decision making, intelligence can be built into enterprise reports to monitor how well they are serving the users. Based on that intelligence, and based on the usage patterns of the reports, even better reports (built by users themselves or by IT) can be constructed, along with the infrastructure that supports those reports. Equally important, reports that do not prove valuable can be discontinued.

At Pace we experienced genuine concerns about the paradigm and practice changes. By design, the intention is to enable end users to build their own reports. A shift from IT-generated reports to end-user-generated reports can be extremely powerful, but there are concerns one needs to contend with. The functional staff feel that this will be an added burden on their already heavy workload, and the IT staff engaged in report delivery feel that their jobs may be in jeopardy as a result. These are very real concerns, and we need to understand that a wholesale shift in responsibilities is not necessarily in the institution’s best interest. Rather, we might wish to leverage the tremendous knowledge and expertise of functional leaders, data warehouse specialists, functional personnel, and software engineers to form collaborations that can efficiently and effectively serve the institution. Smart collaborations will result in a far more efficient data-driven business operation for the institution overall.

**Key Questions to Ask**

- How do we distinguish between “institutional” data (as maintained in enterprise-wide systems) and “local” data (as maintained in departmental databases and shadow systems)? To what degree are our institutional leaders aware of these distinctions?
What are the appropriate uses for data mined from enterprise systems versus data mined from departmental systems? How accurate is the data in these various systems?

Which enterprise-wide reports are the most useful for strategic decision makers? Specifically, which financial reports, enrollment reports, financial aid reports, HR reports, achievement reports, and so forth do we find ourselves using time and again? Which reports do we rarely look at?

What information do we need for decision making that we don’t yet have?

How do we monitor and document which reports are well used and which ones no longer need to be retained?

What model do we use for data access and data distribution? Where are the definitions of this model documented?

Where to Learn More


Acknowledgments

ECAR wishes to thank Mr. Dana Forsberg, enterprise reporting software engineer at eCollege, a Pearson company, for his insightful review and contributions to this research bulletin. Ravishanker wishes to thank Said Reda and Vedika Thacker from Pace
University’s Enterprise Reporting Group for their valuable insights and Dan Venedam from iStrategy Solutions for providing valuable reference material for this work.

Endnotes


About the Author

Ganesan (Ravi) Ravishanker (gravishanker@wellesley.edu) is the Chief Information Officer at Wellesley College. Until recently, he served as Vice President for Information Technology and Chief Information Officer at Pace University.

Citation for This Work


Copyright

Copyright 2011 EDUCAUSE and Ganesan (Ravi) Ravishanker. All rights reserved. This ECAR research bulletin is proprietary and intended for use only by subscribers. Reproduction, or distribution of ECAR research bulletins to those not formally affiliated with the subscribing organization, is strictly prohibited unless prior permission is granted by EDUCAUSE and the author.