Using Dimensional Data to Address Enrollment Management Questions in Higher Education

Enrollment management is critical to the success of colleges and universities. Unfortunately, enrollment management is a difficult task because of the complex set of interrelated questions that needs to be addressed. This article presents many of these questions and explains the process of using dimensional data modeling to design an information system that can provide answers. This article also presents a logical model capable of addressing core issues in enrollment management and explains how to extend it.

By Bob Duniway and Karl Wiegand

Enrollment management is central to the success of a college or university. A school must enroll students into courses, completion of a series of which will lead to graduation. While all colleges and universities must manage the enrollment process at least operationally, it is challenging to keep track of all the related data that would allow for managing the process more strategically.

Recent advances in business intelligence tools can help address the complex topic of enrollment management. Among the most promising advances are software for building dimensional data structures (often referred to as OLAP [On-Line Analytical Processing] cubes) and user-friendly tools for analyzing these structures. Progress in adopting these tools, however, has been slow because enrollment management raises questions that are difficult to organize in a dimensional schema.

In this paper we propose a general framework for applying dimensional modeling to a set of enrollment management questions that are common to many colleges and universities. We begin with an articulation of some standard enrollment management questions grouped by focus. We then use these questions to build a dimensional model around a small set of fact tables. After explaining how the model is derived from these questions, we show how the model can be extended to address related subject areas or questions unique to a particular school.

This paper confines itself to a logical model; implementation of the model depends on many factors, including performance considerations and technology choices.

TYPICAL QUESTIONS

A large number of questions arise in the process of developing, evaluating, and implementing an enrollment management strategy. Although it is beyond the scope of this paper to attempt an exhaustive list of such questions, we nevertheless present—with reference to publications from AACRAO (Westman 2005) and input from participants in the Higher Education Consortium BI/Dashboard group—a list of questions common to a number of different institutions. These questions, organized by subject area, provide the basis for our dimensional data model.

Admissions

One of the biggest issues in enrollment management is that of attracting students to the school. For colleges pursuing enrollment plans that target a certain number, academic quality, or diversity of students, it is crucial to have an appropriate admissions strategy and the ability to monitor that strategy's performance. And it is not only selective colleges that must closely monitor admissions data; colleges with open admission policies also must monitor the number of expected students in order to ensure that instructional resources are sufficient to meet demand.

Some common admissions questions are:

- How many inquiries/prospects do we have by program, college, and student class level?
- How many applicants do we have this year, by program, college, and student class level?
- What percentage of prospects applies? How does this number vary by program, class level, age, gender, race, zip code, and method of contact?

- What percentage of applicants is admitted, by program, class level, age, gender, race, zip code, high school, GPA, and test scores?
- How many admitted applicants accept an offer of admission, by program, class level, age, gender, race, zip code, high school, GPA, test scores, financial aid offer, and financial need?
- What is the average high school GPA and/or class rank of applicants? Of admitted students? Of enrolled students?
- What is the average SAT/ACT score of incoming undergraduate students? Of admitted applicants? For graduate school applicants, what is the average GRE/LSAT/MCAT/...?
- How many applicants do we expect will become new students this year?

Retention and Graduation

Most colleges and universities are less interested in registering students into individual courses than they are in graduating students who complete particular academic programs. Once a group of students is admitted, its retention and graduation patterns can provide an institution with important information about how successfully it is achieving its academic goals.

Some common questions in this area include:

- What percentage of new students completes the first term and returns for the second?
- What percentage of new undergraduates completes a full credit load in the first term? In the first year?
- What is the first-year to second-year retention rate?
- What percentage of students completes the second year and returns for the third?
- What percentage of students entering in a given program graduates in that program? What percent graduates in a different program or major?
- How many students dropped out of a given program this year? What percent transferred to another program or left the college entirely?
- How many students have graduated from a given program each year over the past five years?
- What percentage of students graduates on time (usually two- or four-year rates)?
- What percentage graduates within 150 percent of "on time?"
- What percentage of students ever graduates?
- What is the average number of credits taken by successful graduates?
- What characteristics most distinguish graduates from non-graduates?

Course Section Management

In most colleges, the basic unit of education provided to students is enrollment in a particular course section. Schools must provide the appropriate educational resources to run their scheduled course sections. Schools also must attract sufficient numbers of students to make scheduled course sections economically viable. Managing course section enrollment is essential to supporting the academic progress of students and the economic efficiency of schools.

Some common questions in this area include:

- How many total credit hours of enrollment are there for each course section? For all course sections of a given course, department, course level, or term? How do these numbers compare to past terms?
- How many course sections and course section seats were offered for a given course? By a department at a given course level? Total? How do these numbers compare to past terms?
- What percentage of course section seats was filled? How does this vary by course, department, course level, faculty member, time, and location?
- What percentage of students withdrew from a course section during the term? How does this vary by course, department, course level, faculty member, time, and location?

Tuition Revenue

When students pay tuition, whether they pay per course, per credit hour, or a flat rate for a term, they are purchasing education. Tuition revenue is used to offset instructional costs. Displaying the intersection of tuition revenue and instructional costs is extremely useful for strategic enrollment management.

Some common questions in this area include:

- What is the total tuition revenue generated per term? By department, school/college, and academic level?
- What is the ratio of tuition revenue to faculty cost by course section, course, department, or academic unit?

- After deducting institutional grants, how much does the average student pay per credit hour? By academic level, department, or program?

Faculty Management

Faculty members are the primary providers of the education that students purchase. Faculty salary also is one of the biggest variable costs in providing course sections; consequently, it is particularly useful to track how much it costs to staff a given course section with a given faculty member. In addition to teaching, faculty members also are paid for other work, such as administration and research; this makes cost calculations even more complex.

Some common questions in this area include:

- How many faculty members does the institution have? By department? By academic rank? By demographic characteristics?
- What percentage of the faculty is tenured? On a tenure track? By department? By demographic characteristics?
- What percentage of the faculty is employed full time?
- What are the base salaries of contracted faculty members?
- What are the course section pay rates for adjunct faculty?
- What is the cost to have a given faculty member teach?
- What is the average number of course sections, students, and credit hours being taught by full-time faculty members? By department? By academic rank?
- What is the average number of students in each course taught by a given faculty member?
- What is the percentage of students that drops out of a given faculty member's class?
- How many courses does a given faculty member teach in an academic year? In a particular term?
- What is the overall ratio of students to faculty? By department and program?

Classroom Management

Classroom space limits the number of course sections a higher education institution can offer. Because revenue depends on enrollment in available course sections, and because course sections often require a physical location, it is important for institutions to maximize their use of classroom space. Classroom management involves determining the most efficient use of classrooms to meet course section demands. To be most effective, schools must be aware both of *unmet* course section demand created by classroom shortages and of the inefficiency of underutilized classroom space.

Some common questions in this area include:

- Is our school at capacity? By time slot?
- Which buildings are closest to being used at full capacity?
- Which classroom types are closest to being fully scheduled? Are they utilized by courses that require the specific features of that classroom type? How many courses scheduled in technology-enhanced rooms are using the technology?
- How many more courses of a given type should be scheduled to meet demand?
- What percentage of maximum classroom capacity is being used by actual enrollment?

Total Enrollment Tracking

While it is important to track students by individual course sections, it also is important to track the size and composition of the student population as a whole. Most higher education institutions track enrollment by term, so the model we present in this paper is term-based.

Some common questions in this area include:

- What is our total undergraduate, graduate, and professional enrollment? How does it vary over time?
- What is our enrollment by major and by department?
- What is the demographic composition of our student population?
- How many full-time and part-time students do we enroll? By major, department, and academic level?
- What are the biggest departments by number of majors?
- How many new students enrolled this term? This year? By major, department, and academic level?
- How many students returned this term? This year? By major, department, and academic level?

Financial Aid

To achieve enrollment objectives and to make it possible for students with limited financial resources to complete degree programs, many colleges and universities award financial aid. Knowing how admitted applicants respond to financial aid packages gives schools the opportunity to use aid strategically to achieve enrollment objectives. Knowing how aid affects student retention and graduation also helps schools evaluate whether the aid is sufficient to support student success. Finally, monitoring the distribution of institutional grant aid (also known as "discounting") ensures that schools have accurate information regarding the net tuition revenue received from students. A degree program that is populated with students receiving no discount will generate more net tuition revenue than a similarly sized degree program populated with students receiving substantial institutional grants.

Some common questions in this area include:

- What is the average discount rate for all students; for new students; for students in a particular program; or for students with certain characteristics?
- How much did the average discount rate of undergraduates change from last year? Of graduate students? By program and major?
- What is the average financial need of our students? How has that changed over time?
- How does the yield of admitted applicants vary by financial need? By unmet financial need? By institutional grant amount?
- What was the average unmet financial need after financial aid?

FACTS AND CATEGORIES

With so many interrelated questions, we need a unified approach that can improve the consistency of reporting and increase the efficiency of our data management and allow us to explore interdependencies. By relying on analysis rather than trusting untested assumptions, schools can implement better strategies for enrollment management.

A dimensional model achieves this result by collecting numeric measures in one or more fact tables and linking these measures to the various characteristics of the event generating the fact(s) by linking to a set of dimension tables. *Facts* are numbers that can be added and averaged for analysis. *Dimension attributes* are used to group facts so that sums and averages for groups are meaningful. For the sake of clarity, this paper ends table names with "…Fact" for fact tables and "…Dim" for dimension tables.

For example, a student registered in a course section has *facts* such as credit hours attempted, credit hours earned, gross tuition, and net tuition; these might be placed in a CourseSectionFact table. All of the characteristics of the student would be stored on a student dimension table while all the course section characteristics would be stored on a course section dimension table. Analyzing the questions listed above, it is possible to create two dimensional models which can encompass the full complexity of enrollment management.

Admissions

In the area of admissions, the central facts appear to be counts of applicants and counts of what happens to those applicants. Those counts are divided further according to the personal characteristics of the applicants and the programs to which they apply.

To address the level of detail in the gathered questions, our model stores facts at the level of one potential applicant to one program for one term in StudentProgramCareerFact. Our facts for this record reflect the achievement of each stage of the admissions process, *e.g.*, applied, was admitted, etc. We represent this as a set of boolean variables set to "1" when the individual has reached a particular stage, such as "admitted," or "0" if the applicant has not yet reached that stage. We also store as a fact the date when a student attains a stage of the admissions process.

Our logical model (see Figure 1, on page 37) treats information about the individual as attributes on a PersonDim table. To address questions at the program level, we added a ProgramDim table which allowed us to group applicants by program and to sum individual facts to get counts by program. Additional questions require grouping these facts based on application characteristics, such as the department or college to which the applicant is seeking admittance and the applicant's financial aid status. In order to accommodate these characteristics, we needed additional dimension tables, such as ApplicationDim and FinancialAidAppDim.

Retention and Graduation Rates

Examination of questions in the area of retention and graduation rates suggested that we needed to stay at the level of tracking individual student progress through programs. We simply extended the first model to include additional

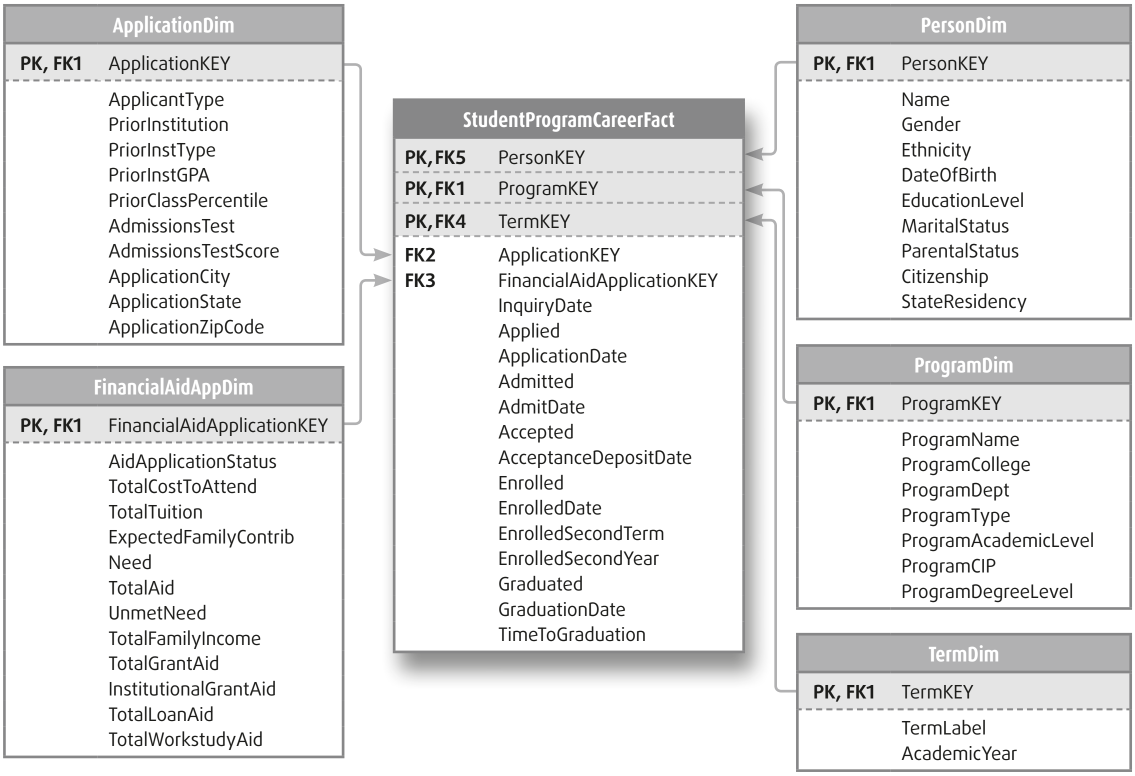

FIGURE 1. *Longitudinal Enrollment Management Logical Model*

boolean and date facts: returning the second term, returning the second year, graduation date, and graduating from the program and/or college to which they initially gained admittance.

Course Section Management

When we examined questions about course section management, we found that our initial fact table was no longer suited to the facts we needed to track. We still wanted to address questions at the student/program/term level, but we also were interested in facts about students enrolled in course sections. Thus, our focus shifted from a longitudinal view of a student's career to a cross-sectional view of a student's enrollment behavior in a particular academic term (see Figure 2, on page 38). For these questions, StudentCourseSectionFact contains one record for each student registration in a particular course section; facts include the number of credits attempted, the number of credits earned, and whether the student withdrew from the course section.

In order to fully address the questions in this section, however, it proved insufficient to simply add up the student enrollment facts. We therefore added CourseSectionFact, which contains information about the course section (such as section, room capacity, and faculty member) and is linked to StudentCourseSectionFact via a CourseSectionKey.

Tuition Revenue

Questions concerning tuition revenue proved a particular challenge. We wanted to monitor tuition revenue in a flexible manner, which would have been relatively easy if we had associated tuition as a fact of each student's course section enrollment; yet many programs charge a flat tuition rate *per term* rather than per course or per credit. To calculate tuition revenue, we therefore needed a StudentTermFact table that included the total term tuition along with the total credits attempted. With the addition of this table, calculating per-credit tuition became a simple matter of division.

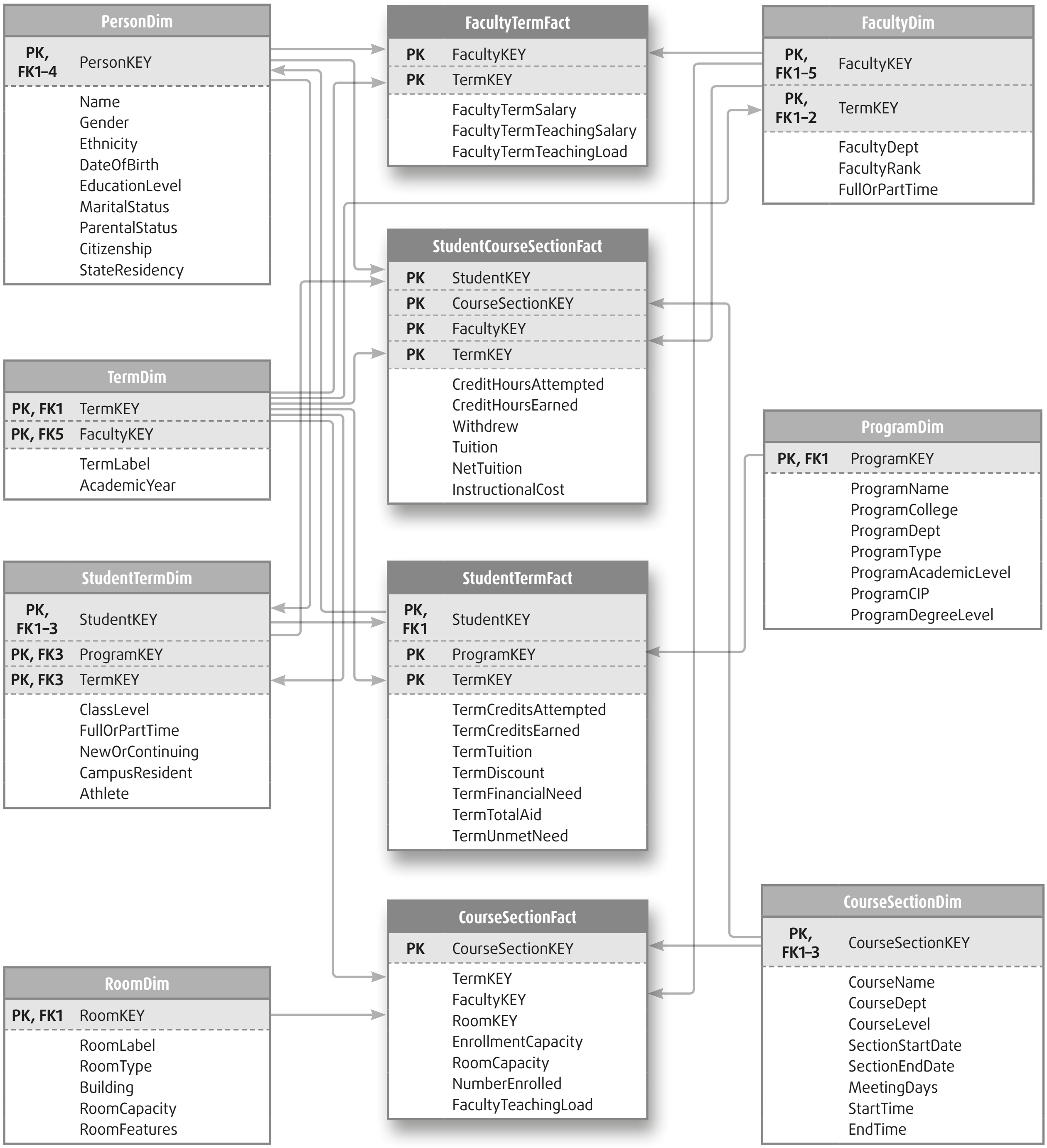

FIGURE 2. *Cross-Sectional Enrollment Management Logical Model*

NOTE: In this model, StudentKey and FacultyKey are synonyms for PersonKey.

The advantage of this approach became evident when we saw that in addition to face-value tuition, there were common questions about net tuition after institutional grants were deducted. The StudentTermFact table became a place to track total term institutional grant aid.

Whether this approach works for any given institution depends on exactly how tuition (and other instructional revenue sources, such as course fees or government-provided instructional funding) relate to the fact of a student being enrolled in a particular course and/or for a particular term in a particular program. In moving from logical model to implementation, this issue will need to be carefully resolved by each individual institution.

Faculty Management

FacultyTermFact contains faculty salary and teaching load data for the term. These facts allow us to calculate instructional cost at the course-section level by calculating the percentage of the total instructional load represented by the section and then multiplying that by the term teaching salary.

FacultyDim contains information about each faculty member, such as department, rank, and full- or part-time status. These characteristics may change from term to term, so each record of this table is defined at the faculty-member-per-term level of granularity.

Classroom Management

In addressing questions related to classroom management, we saw the value of constructing a comprehensive model. All we needed to add to our existing model were a room dimension connected to the CourseSectionFact table and a room capacity fact on that table.

Total Enrollment and Financial Aid

Questions relating to total enrollment and financial aid did not require us to significantly alter our model. Data already in StudentCourseSectionFact and StudentTermFact can be summed to determine aggregate enrollment and grant aid. However, these questions involve some term-dependent aspects of a student's relationship with the school, such as new or continuing status, so we added a StudentTermDim table to hold these attributes. Financial aid questions require a number of additional facts about students; these facts were added to the StudentTermFact table. The model likely will need to be extended at each institution by including in the StudentTermFact table other aid measures of interest.

EXTENDING OR MODIFYING THE MODEL

Flexibility is one of the many advantages of the model presented in this paper. Many schools will look at this model and recognize that it does not address a critical enrollment management issue on their campus. However, as long as students pursue completion of academic programs longitudinally and enroll in course sections within terms, it is likely that the model can be modified to account for local concerns. For example, if an institution is focused on academic performance as measured by grade point average, the model can be modified to include grade information. If two facts, GPA_Points and GPA_Credits, are added to the StudentCourseSectionFact table, the formula GPA = Sum(GPA_Points)/Sum(GPA_Credits) can be applied to any collection of course section grades. The same formula can be used to calculate average grades given in a course section; to compare average grades of students by personal characteristics; and to compare the average grades given by academic departments.

Suppose a particular institution wants to evaluate a group of programs—such as natural science programs—collectively. It could add an element to the ProgramDim table, label it something like ProgramCategory, and proceed with analysis of that grouping.

A slightly more difficult modification could be made to analyze the impact of various interventions on retention and graduation. The desired information could be tracked by adding an InterventionDim table to the longitudinal model, connecting it to the StudentProgramCareerFact table with a new InterventionKey, and then adding fields to describe participation in the various interventions being evaluated.

When attempting to obtain new functionality, it often is faster and easier to modify an existing model than to create a new one. Even in cases where it might be faster to build a new model, there are advantages to extending the existing model. Because of the number of interrelated factors available, existing facts and dimensions will support flexible analysis, leading to deeper understanding of critical issues.

ANALYSIS ACROSS THE MODELS

Some questions may require investigation of the interaction between the longitudinal and cross-sectional models. For example, an institution may want to know if poor performance in particular courses results in students' leaving a particular program or even the school. The sharing of dimensions between these two models makes it possible to develop new data structures to perform such analysis. The PersonDim, ProgramDim, and TermDim tables make it easy to identify the first-term courses taken by new students in any program. To analyze the retention impact of students' performance in a particular course, the analyst would need to add measures of performance such as EnrolledInProgramSecondYear and GraduatedInProgram to the StudentProgramCareerFact table.

IMPLEMENTING THE MODEL

Modifying the model to address all of the enrollment management issues at a particular institution likely will lead to a significantly more complex model than that presented in this paper. Nevertheless, the process of examining key questions and developing a logical model to address them is both informative and useful. Developing a full logical model does not necessitate full implementation. Implementations can be staged.

Seeing the full model in advance enables design and implementation of system modules that eventually will work well together. Knowing that admissions questions are logically related to questions of retention and graduation could lead to the inclusion of dimensions and dimension attributes that will prove extremely useful when the time comes for retention and graduation analysis. If the logical model is developed using a tool such as Microsoft SQL Server Analysis Services or Microsoft Visio, the modeling process can transition seamlessly from explaining the logical model to providing documentation for technical staff. With such a tool, staff can begin building a physical implementation of the model with confidence that the design has been clearly specified. Documenting the metadata as the model is implemented becomes a straightforward process, resulting in a system that is easier to maintain, explain, and extend.

When implementing a dimensional data system for enrollment management, it is important to remember that tools available to users usually show only a partial view of the entire database. A partial view is convenient because the hiding of unnecessary fields lessens confusion for beginning users. However, as users become more advanced and/or as it becomes necessary to include additional information for purposes of analysis, it is easier to take that information into account if it already is in the database.

Moreover, the challenges of data extraction, transformation, and loading (ETL) into the dimensional data mart can be substantial. By developing the logical model first, the ETL process can be measured against a clearly defined information objective, and the impact of data limitations can be readily understood in terms of the questions being addressed. It might be a poor decision to restrict the logical model simply because it is difficult to implement given currently available data and development resources. Also, it is not unusual for the process of developing an ETL strategy to expose additional data that address unanticipated questions and provide real value, requiring extension of the logical model.

DATA DEFINITIONS

Moving from the logical model to implementation of the physical model will require precise definition of each data element. Even common terms such as "applicant," "new student," and "aid" are subject to different interpretations by different users. The term "freshman" may refer to any undergraduate who hasn't yet reached sophomore standing, or it may refer to the entering cohort of new, first-time, full-time undergraduates. These distinct student statuses need both to be tracked differently in the database and to be given distinct labels in all reports; only then can confusion be avoided.

To support benchmarking against other institutions, it may be worth using definitions from a standards body such as the Integrated Post-secondary Educational Data System (IPEDS) or the common data set (CDS). Unfortunately, standards probably won't address all data elements needed for enrollment management. Use standards when they are available; clearly document how each data element is defined; give data elements unambiguous names; and educate users as to data definitions.

CONCLUSION

Because of the large number of related questions raised in the course of developing and implementing an enrollment strategy, the process of developing a comprehensive data model can be daunting. However, it is precisely because of this complexity that a systematic approach to supporting enrollment management analysis and reporting is worthwhile. Making informed decisions in this area frequently calls for answering a series of questions about the likely impact of changing one aspect of the system. A dimensional data mart that accounts for these overlapping relationships in a carefully designed logical model lowers the cost of exploring these questions, yielding more informed conversations and more successful decisions.

REFERENCES

Bontrager. B. 2004. Enrollment management: An introduction to concepts and structures. *College and University*. 79(3).

Loshin, D. 2007. Requirements for modeling metrics. *b-eye-network.com*. Sept. 20. Boulder, CO: BeyeNETWORK. Retrieved from: <www.b-eye-network.com/channels/1148/view/6026>.

Luan, M., and C. Zhao. 2006. *Data Mining in Action: Case Studies of Enrollment Management: New Directions for Institutional Research*. San Francisco: Jossey-Bass

Maltz, E.N., K. E Murphy, and M.L. Hand. 2007. Decision support for university enrollment management: Implementation and experience. *Decision Support Systems*. 44(1): 106–123. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S016792360700053X>.

Westman, C., and P. Bouman. 2005. *AACRAO's Basic Guide to Enrollment Management.* Washington, D.C.: American Association of Collegiate Registrars and Admissions Officers.

About the Authors

DR. BOB DUNIWAY, as Director of Institutional Research at Seattle University, has spearheaded the in-house development of a comprehensive dimensional data warehouse and is a key participant in strategic enrollment management. He has spoken on the use of dimensional data in higher education at HEDW (2007 and 2008), EDUCAUSE (2008) and the Campus Technology Winter Conference (2007). He is author of *Winning the College Game: How to Find, Get into, Pay for and Thrive in a Great College* (2008).

KARL WIEGAND is a Ph.D. candidate in computer science at Northeastern University, where he is focusing on natural language processing and machine learning. He completed his M.S. in Software Engineering at Seattle University, where he was awarded a grant from Microsoft to do work on business intelligence in higher education; this work was foundational to developing the model presented in this article.

ACKNOWLEDGMENTS

The authors are indebted to members of the Microsoft Higher Ed Consortium (HEC) for proposing this paper (www.microsoft.com/education/hec), for the support and encouragement of Adrian Wilson and Jim Ptaszynski at Microsoft Higher Education, and for invaluable feedback from the HEC Dashboard/BI Council. We are particularly grateful for the extensive feedback and suggestions provided by Christopher Brown, St. Cloud State University; Mike Conlon, University of Florida; John Gohsman, University of Michigan; and Keith Mutch, Bellevue Community College.